Jon Elster

COMPORTAMENTUL SOCIAL

Fundamentele explicației în științele sociale

Cuprins

PREFAȚĂ 11
INTRODUCERE 15

I ⇝ EXPLICAȚIE ȘI MECANISME 21
1. Explicația 23
2. Mecanismele 43
3. Interpretarea 61

II ⇝ GÂNDIREA 75
4. Motivațiile 82
5. Interesul personal și altruismul 100
6. Miopie și hipermetropie 114
7. Credințele 126
8. Emoțiile 145

III ⇝ ACȚIUNEA 161
9. Dorințe și oportunități 163
10. Persoane și situații 175
11. Alegerea rațională 187
12. Raționalitate și comportament 207
13. Reacții la iraționalitate 224
14. Câteva implicații asupra interpretării textelor 236

IV ⇝ LECȚII DESPRINSE DIN ȘTIINȚELE NATURII 247
15. Fiziologie și neuroștiințe 251
16. Explicația prin consecințe și selecția naturală 260
17. Selecția și comportamentul uman 274

BCU IAȘI - CENTRAL UNIVERSITY LIBRARY

V ❧ INTERACȚIUNEA 285
18. Consecințe neintenționate 286
19. Interacțiunea strategică 297
20. Jocurile și comportamentul 313
21. Încrederea 324
22. Normele sociale 332
23. Formarea de credințe colective 348
24. Acțiunea colectivă 362
25. Decizia colectivă 373
26. Organizații și instituții 396

CONCLUZIE: ESTE POSIBILĂ ȘTIINȚA SOCIALĂ? 411

INDEX 433

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Cartea de față este o ediție extinsă și revizuită a volumului cu mare succes de critică *Nuts and Bolts for the Social Sciences*. Jon Elster abordează în 26 de capitole succinte natura explicației în științele sociale oferind, în același timp, o analiză a stărilor mentale – credințe, dorințe și emoții – care sunt precursoare ale acțiunii, o comparație sistematică a modelului comportamental fundamentat pe teoria alegerii raționale cu sistemele alternative, o discuție cu privire la ce pot învăța științele sociale din neuroștiințe și biologia evoluționistă, precum și o trecere în revistă a mecanismelor interacțiunii sociale, de la comportamentul strategic până la decizia colectivă. Autorul oferă o prezentare generală a unor mecanisme explicative-cheie în științele sociale, bazându-se pe sute de exemple și făcând apel la o mare varietate de surse – psihologie, economie comportamentală, biologie, științe politice, istorie, filosofie și literatură. Folosind un limbaj accesibil și eliberat de jargonul specializat, Elster încearcă să fie cât mai precis și clar în expunerea sa, fără să recurgă, însă, la modele formale. Într-o concluzie provocatoare, Elster apără caracterul central al științelor sociale calitative atât împotriva obscurantismului slab (literar), cât și a obscurantismului tare (de tip matematic).

Jon Elster este Profesor (Chaire de Rationalité et Sciences Sociales) la Collège de France. Fellow al Academiei Americane de Arte și Științe, precum și Corresponding Fellow al Academiei Britanice, este beneficiarul unor burse acordate de Fundația John Simon Guggenheim și de Fundația Russell Sage, printre altele. Dr. Elster a predat la Universitatea din Chicago și la Universitatea Columbia și a deținut calitatea de profesor-invitat al multor universități din Statele Unite și Europa. Este autorul sau editorul a treizeci și patru de cărți, printre cele mai recente numărându-se: *Transitional Justice in Historical Perspective* și *Retribution and Restitution in the Transition to Democracy*.

BCU IAȘI - CENTRAL UNIVERSITY LIBRARY

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Pentru Jonathan și Joanna

B.C.U. „M. EMINESCU" IAȘI

BCU IASI/CENTRAL UNIVERSITY LIBRARY

BCU IASI/CENTRAL UNIVERSITY LIBRARY

PREFAȚĂ

Am început să scriu cartea de față cu intenția de a revizui o carte pe care am publicat-o în 1989, *Nuts and Bolts for the Social Sciences*. Într-un final, a ieșit o carte destul de diferită și mult mai ambițioasă decât am dorit inițial. Cartea acoperă o varietate mult mai mare de subiecte, cu o cantitate semnificativ mai mare de detalii și într-un mod de abordare diferit. Cu toate că nouă capitole au aceleași titluri ca în cartea precedentă, doar capitolele 9 și 24 au rămas în bună măsură nerevizuite.

Deși este cuprinzătoare, cartea nu se vrea un tratat. Este, în același timp, mai puțin, dar și mai mult decât un tratat. Este, într-un mod elementar, informal și personal, o prezentare de idei care au, cred eu, un potențial considerabil pentru a lămuri comportamentul social. Folosesc foarte multe exemple, multe dintre ele cu caracter anecdotic sau literar, altele făcând apel la studii mai sistematice. Algebra, utilizată în mod sporadic, nu depășește nivelul de liceu. În același timp, cartea are o abordare metodologică și filosofică rar întâlnită în cărțile de nivel introductiv. Se conturează o încercare de a plasa științele sociale în domeniul mai larg al științelor în general, domeniu care include atât științele naturale, cât și științele umaniste. Se conturează, de asemenea, o încercare de a-l face pe cititor conștient de modul în care principiile generale ale explicației reglementează construirea de teorii cu pretenții explicative.

Caracterul notelor bibliografice oferite pentru fiecare capitol reflectă importanța din ce în ce mai mare a Internetului, în special a unor surse ca Wikipedia, Google.com și Scholar.Google.com. Întrucât cititorii pot găsi referințele relevante în timp foarte scurt, am omis specificarea surselor pentru multe dintre afirmațiile pe care le fac și pentru descoperirile prezentate. În schimb, încerc să-i ghidez pe cititori către cărți-sursă importante, către unele dintre cărțile moderne clasice, către cărți și articole de specialitate care sunt sursele unor teze mai greu de depistat cu ajutorul Internetului și către autori de la care am împrumutat atât de mult încât, dacă nu i-aș menționa,

BCU ASCR CENTRAL UNIVERSITY LIBRARY

acest lucru ar justifica un joc de cuvinte cu propriul meu nume (*Elster* în germană înseamnă coțofană).[1]

Cu toate că textul propriu-zis conține puține referințe la autori contemporani, îi menționez *in extenso* pe Aristotel, Seneca, Montaigne, La Rochefoucauld, Samuel Johnson, H.C. Andersen, Stendhal, Tocqueville, Proust, precum și pe alți autori clasici, care rămân surse literare inepuizabile de ipoteze cauzale. Ne-am debarasa de multe intuiții prețioase dacă am ignora mecanismele sugerate de filosofie, literatură, teatru și poezie. Dacă neglijăm douăzeci și cinci de secole de meditație asupra minții, acțiunii și interacțiunii în favoarea ultimei sute de ani sau a ultimilor zece ani, atunci ne expunem pericolului de a pierde foarte mult. Îi citez pe acești autori nu atât pentru a recurge la autoritatea lor, ci pentru a da motive temeinice în favoarea unei lecturi cuprinzătoare, mult mai valoroasă decât lectura de nișă. Fac aceasta în opoziție directă cu ceea ce mi se pare că este o continuă profesionalizare a științelor sociale (în special în America), tendință care nu-i încurajează deloc pe studenți să învețe limbi străine și să citească literatură mai veche; volumul de față este o pledoarie extinsă pentru o abordare cuprinzătoare a studiului societății.

În pregătirea manuscrisului, am beneficiat de asistență și comentarii din partea multor persoane. Ar trebui să mulțumesc, în primul rând, studenților mei de la Universitatea Columbia pentru întrebările lor agere și pentru comentariile emise la cursul în care am prezentat pentru prima dată ideile expuse în cartea de față. Sugestiile venite de la Pablo Kalmanovitz mi-au fost în mod deosebit folositoare. Aanund Hylland și Ole-Jørgen Skog au petrecut trei zile cu mine la Collioure discutând o primă versiune a cărții. Apoi, în Oslo, împreună cu Hyllard, Karl O. Moene și John Roemer, am continuat discuția pe parcursul unei zile și jumătate. Comentariile lor nu numai că m-au salvat de la multe (multe!) greșeli, dar mi-au sugerat, de asemenea, modul în care aș putea să completez și să consolidez expunerea ideilor. Îi sunt recunoscător în special lui Roemer pentru faptul că m-a îndemnat să scriu o concluzie. Am primit comentarii referitoare la întregul manuscris de la Diego Gambetta, Raj Saah și de la un recenzent care a preferat să rămână anonim. Comentariile lui Gambetta merită menționate în mod special întrucât au fost foarte detaliate și de mare ajutor. Am purtat conversații utile cu Walter Mischel despre ideile prezentate în capitolul 10 și care, în bună măsură, îi aparțin. Am primit și comentarii în scris, foarte valoroase, de la George Ainslie cu privire la ideile prezentate în Partea I a cărții, multe dintre acestea aparținându-i. Bernard Manin mi-a oferit comentarii constructive asupra capitolului 25. Robyn Dawes m-a ajutat cu comentarii foarte pertinente asupra

[1] Autorul se referă la celebra operă a lui Gioacchino Rossini *La gazza ladra – Coțofana hoață*. (N. tr.)

capitolelor 7 și 12. În cele din urmă, de-a lungul ultimilor șapte ani, am prezentat versiuni preliminare ale capitolelor din cartea de față membrilor „Grupului de luni", care a avut întâlniri săptămânale în New York în fiecare toamnă și, ocazional, primăvara, din 1995 încoace: John Ferejohn, Raquel Fernandez, Russell Hardin, Stephen Holmes, Steven Lukes, Bernard Manin, Pasquale Pasquino, Adam Przeworski și John Roemer. Le mulțumesc tuturor pentru obiecțiile constructive și prietenoase.

Dedic această carte lui Jonathan și Joannei Cole – vor înțelege ei de ce.

Citez *Eseurile* lui Montaigne în traducerea lui M. Screech (Londra: Penguin, 1971); referințele la Proust provin din noua traducere editată de C. Pendergast (Londra: Penguin, 2003); *Cugetările* lui Pascal în traducerea lui A. J. Krailsheimer (Londra: Penguin, 1995); *Maximele* lui La Rochefoucauld în traducerea lui L. Tancock (Londra: Penguin, 1981); *Caracterele* lui La Bruyère în traducerea lui H. van Laun (New York: Scribner, 1885); *Despre iubire* a lui Stendhal în traducerea lui G. Sale, S. Sale și J. Stewart (Londra: Penguin, 1975); și *Despre Democrație în America* a lui Tocqueville în noua traducere a lui A. Goldhammer (New York: Library of America, 2004). Celelalte traduceri din franceză îmi aparțin.[1]

[1] Pentru traducerea românească, trimitem cititorii la: Michel de Montaigne, *Eseuri*, Editura Științifică, București, 1966, Marcel Proust, *Opere*, 3 vol., Editura Academiei Române, București, 2011, Blaise Pascal, *Cugetări*, în *Scrieri alese*, Editura Științifică, București, 1967, La Rouchefoucauld, *Maxime și reflecții*, Editura Minerva, 1967, La Bruyère, *Caracterele*, 2 vol., Editura pentru Literatură Universală, București, 1968, Stendhal, *Despre dragoste*, Editura pentru Literatură Universală, București, 1968, Alexis de Tocqueville, *Despre democrație în America*, 2 vol., Humanitas, București, 2005. (N. red.)

BCU MASIV CENTRAL UNIVERSITY LIBRARY

BCU IASI/CENTRAL UNIVERSITY LIBRARY

INTRODUCERE

Cartea de față are în vedere explicarea comportamentului social. În prima parte prezint într-o formă detaliată concepția pe care o am despre explicație, iar în celelalte patru părți, construiesc instrumentarul de concepte și mecanisme care se aplică în cazuri particulare. Nu cred că mai este nevoie să precizez că nu urmăresc să prezint un tablou complet. În loc să încerc o prezentare în detaliu a lacunelor, care vor fi evidente, voi începe prin a enumera câteva exemple de probleme paradoxale care pot fi lămurite prin abordarea mea. În *Concluzie* mă voi întoarce la aceste situații, pe care le voi completa cu scurte referințe la explicațiile pe care le citez în capitolele anterioare.

Exemplele care urmează și explicațiile oferite trebuie luate în considerare ținând cont de două avertismente. În primul rând, nu susțin că toate lucrurile care trebuie explicate sunt fapte bine-cunoscute. În cazul unor explicații concrete, acest lucru este un prim pas crucial – nu are sens să cauți o explicație pentru ceva ce nu există. Însă în scopul elaborării unui instrumentar, putem fi mai puțin riguroși. În al doilea rând, chiar și pentru lucrurile care trebuie explicate (*explananda*) și al căror caracter de fapt este bine documentat, nu pretind că explicațiile pe care le menționez sunt corecte. Pretind doar că acestea satisfac o condiție minimală pentru a fi considerate explicații – aceea că implică din punct de vedere logic lucrurile care trebuie explicate. Problemele și explicațiile pe care le voi prezenta au scopul de a arăta că, „dacă un anumit tip de lucruri se întâmplă, atunci acesta este tipul de mecanism care ar putea constitui o explicație", precum și că, „dacă acest mecanism este funcțional, acesta este tipul de lucruri pe care le poate produce". Date fiind aceste două avertismente, putem trece mai departe la prezentarea problemelor paradoxale, organizate într-un mod oarecum arbitrar (întrucât multe dintre aceste situații pot fi încadrate în mai multe categorii), în conformitate cu cele patru secțiuni mari ale cărții.[1]

[1] Deși situațiile din lista de față se intersectează destul de mult cu situațiile prezentate în capitolul 12 drept provocări la adresa teoriei alegerii raționale, nu intenționez aici să stârnesc polemici, ci doar să stârnesc curiozitatea cititorului.

BCU CLUJ CENTRAL UNIVERSITY LIBRARY

I ↝ Gândirea

- De ce cred unii pariori că, după o serie de cinci apariții consecutive de roșu, este mai probabilă apariția roșului decât a negrului?
- De ce cred alți pariori că, după o serie de cinci apariții consecutive de roșu, este mai probabilă apariția negrului decât a roșului?
- De ce, uneori, ni se schimbă preferințele prin simpla trecere a timpului?
- De ce mulți dintre cei care par a crede în viața de apoi vor ca aceasta să survină cât mai târziu posibil?
- De ce sunt oamenii reticenți, față de sine și față de ceilalți, în a recunoaște că sunt invidioși?
- De ce sunt oamenii reticenți, față de sine și față de ceilalți, în a recunoaște că sunt ignoranți în privința un anumit lucru?
- De ce în rândul convertiților la calvinism din secolul al XVI-lea credința că oamenii sunt predestinați să ajungă fie în Iad, fie în Rai aducea mai multă împăcare sufletească decât credința că se poate atinge mântuirea prin fapte bune?
- De ce este (uneori) adevărat că „Cel care face rău nu poate să ierte"?
- De ce este rușinea mai importantă decât vina în anumite culturi?
- De ce victoria Franței la Campionatul Mondial de Fotbal din 1998 a produs atât de multă bucurie în țară și de ce faptul că echipa nu s-a calificat din grupe în 2002 a produs atât de multă mâhnire?
- De ce unele femei trăiesc un sentiment de rușine în urma unui viol?
- De ce ritualurile de inițiere umilitoare sporesc loialitatea față de grupul în care are loc inițierea?

II ↝ Acțiunea

- De ce, în zilele noastre, publicul aplaudă în picioare spectacolele de pe Broadway mai frecvent decât acum douăzeci de ani?
- De ce pedepsele, în loc să reducă rata de incidență a comportamentului pe care îl vizează, mai degrabă o sporesc?

BCU ALEXANDRU IOAN CUZA CENTRAL UNIVERSITY LIBRARY

- De ce sunt oamenii refractari la încălcarea regulilor autoimpuse, chiar și atunci când respectarea lor nu are sens?
- De ce răzbunarea urmează mai degrabă modelul „Doi ochi pentru un ochi" decât „Un ochi pentru un ochi"?
- De ce randamentul pe termen lung al acțiunilor este mai mare decât al obligațiunilor (adică de ce valoarea acțiunilor nu crește, astfel încât să egaleze randamentele)?
- De ce rata sinuciderilor scade atunci când medicamentele periculoase se vând în folii, față cu vânzarea lor în flacoane?
- De ce niciunul dintre cei treizeci și opt de martori care au asistat la uciderea în bătaie a lui Kitty Genovese nu a chemat poliția?
- De ce unii oameni au adăpostit sau salvat evrei în timpul regimului nazist?
- De ce a vrut președintele Chirac alegeri anticipate în 1997, având în vedere că pierdea majoritatea parlamentară?
- De ce unii părinți divorțați sunt dispuși să împartă custodia copilului în ciuda faptului că ar prefera custodia unică, pe care ar obține-o dacă ar merge la tribunal?
- De ce este mai puțin probabil ca oamenii săraci să emigreze?
- De ce unii oameni economisesc în conturi speciale pentru Crăciun, conturi care nu oferă nici dobândă și nici posibilitatea efectuării unei retrageri înainte de Crăciun?
- De ce se angajează oamenii în proiecte cu valoare estimată ca negativă, cum ar fi construcția aeronavei *Concorde*?
- De ce în cadrul „justiției de tranziție" (care presupune că factorii responsabili pentru un regim autocrat sunt aduși în fața unui tribunal după trecerea la democrație) cei judecați mai devreme primesc pedepse mai severe față de cei judecați mai târziu?
- De ce, în piesa omonimă a lui Shakespeare, Hamlet își amână răzbunarea până în ultimul act?

BCU IASI / CENTRAL UNIVERSITY LIBRARY

III ⇝ Lecții desprinse din științele naturii

- De ce este mai probabil ca părinții să își ucidă copiii adoptați sau vitregi, și nu copiii lor biologici?
- De ce este atât de rar întâlnit incestul între frați, în condițiile în care tentația există, la fel și oportunitatea?
- De ce sunt oamenii dispuși să investească în proiectele altora, în condițiile în care cei din urmă păstrează întregul profit?
- De ce uneori oamenii se răzbună suportând costuri materiale, dar fără niciun beneficiu material?
- De ce trag oamenii concluzii care nu sunt susținute de dovezi?

IV ⇝ Interacțiunea

- De ce susținătorii unui partid socialist votează uneori cu comuniștii, împiedicându-și astfel propriul partid să câștige?
- De ce unele state care abia și-au câștigat independența adoptă ca limbă oficială limba foștilor asupritori imperialiști?
- De ce standurile de înghețată sunt situate de cele mai multe ori unul lângă altul în mijlocul plajei, în condițiile în care, dacă ar fi răspândite pe toată plaja, ar fi mult mai bine pentru consumatori, iar pentru vânzători nu ar fi deloc mai rău?
- De ce își exprimă unii oameni opțiunea de a vota chiar și atunci când aceasta nu poate produce niciun efect în privința rezultatului de la urne?
- De ce indivizii de succes din societățile occidentale au, de regulă, o greutate mai mică decât restul oamenilor?
- De ce oamenii se abțin de la tranzacții care ar fi în beneficiul tuturor, ca, de exemplu, faptul de a se abține să întrebe pe cineva aflat în față la coadă dacă nu ar fi dispus să își vândă locul?
- De ce președintele Nixon a încercat să își creeze în fața sovieticilor o imagine de om înclinat spre un comportament irațional?
- De ce comandanții militari incendiază uneori podurile aflate în subordinea lor (sau propriile nave)?

BCU CLUJ-NAPOCA UNIVERSITY LIBRARY

- De ce sunt atât de importante chestiunile care țin de etichetă, în sine nesemnificative?
- De ce oamenii lasă bacșiș taximetristului sau chelnerului chiar și atunci când se află într-un oraș pe care, probabil, nu îl vor mai vizita vreodată?
- De ce firmele investesc în stocuri masive chiar și atunci când nu anticipează o întrerupere a producției?
- De ce fiecare student dintr-un grup are impresia că ceilalți au înțeles mai bine decât el un text obscur?
- De ce în multe adunări legislative se votează prin apel nominal?
- De ce sunt favorurile reciproce mai des întâlnite în adunările legislative ordinare decât în adunările constituante?

În diverse locuri din carte voi propune anumite explicații pentru fenomenele de mai sus și voi oferi și un rezumat în concluzii. Pentru moment, vreau să fac o remarcă generică despre două tipuri de explicație care, probabil, nu vor fi de niciun folos. După cum se va vedea din chiar primul capitol, lucru pe care îl voi puncta și pe parcurs, unul dintre scopurile cărții este acela de a inculca o doză de scepticism față de două moduri de gândire obișnuite. În primul rând, și cu foarte puține excepții, științele sociale nu se pot baza pe explicația funcțională, care dă seamă de acțiuni și caracteristici comportamentale enumerând consecințele acestora, și nu cauzele. Oare normele care reglementează bacșișul există întrucât este mai eficientă monitorizarea chelnerilor de către client, decât de către patron? Eu nu cred. În al doilea rând, cred că teoria alegerii raționale are o putere explicativă mai mică decât obișnuiam să cred mai demult. Oare oamenii concreți acționează în baza calculelor care umplu pagini întregi de anexe formale în cele mai titrate reviste de specialitate? Nu cred.

Cu toate acestea, din cel puțin trei motive, teoria alegerii raționale este o parte valoroasă a instrumentarului. Dacă o vom înțelege într-un sens calitativ, orientat de simțul comun, ea poate să explice destul de mult din comportamentul de zi cu zi. Chiar și atunci când nu explică prea mult, poate avea o valoare conceptuală imensă. În particular, teoria jocurilor a făcut lumină în structura interacțiunilor sociale într-un mod care depășește considerabil înțelegerea dobândită în secolele trecute. Nu în ultimul rând, oamenii *vor* să fie raționali. Dorința de a avea motive temeinice pentru un anumit comportament și de a nu fi doar o marionetă a forțelor psihice care acționează „fără ca noi să știm" reprezintă o forță de opoziție în raport cu mecanismele generatoare de irațional pe care urmează să le trec în revistă.

Cu toate că am o atitudine critică față de multe dintre explicațiile alegerii raționale, cred că noțiunea de *alegere* este fundamentală. În această carte voi lua în considerare mai multe alternative la explicația bazată pe alegere și voi trage concluzia că, deși acestea pot fi de multe ori un adaos util la cea din urmă, nu o pot niciodată înlocui. Faptul că oamenii acționează sub diverse *constrângeri*, de exemplu, poate oferi o explicație pentru o variație substanțială în comportament. De asemenea, în unele cazuri se poate susține că *selecția agenților* este responsabilă pentru comportamentul observat, și nu *alegerea de către agenți*. Însă, în general, cred că factorul subiectiv al alegerii are o valoare explicativă mai mare decât factorii obiectivi ai constrângerii și selecției. Aceasta este, în mod evident, o intuiție care nu poate fi dovedită riguros și, în orice caz, savanții din științele sociale ar trebui să facă loc tuturor factorilor în instrumentarul lor.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

I
EXPLICAȚIE ȘI MECANISME

Cartea de față se bazează pe o perspectivă bine definită asupra explicației în științele sociale. Deși nu este un demers preocupat înainte de toate de filosofia științelor sociale, face apel și susține anumite idei metodologice despre cum anume trebuie explicate fenomenele sociale. Aceste idei sunt expuse în primele trei capitole. În restul cărții, ele devin fundalul ideatic implicit, deși revin în atenție într-un mod explicit pe alocuri, în capitolele 14 și 17, precum și în concluzii.

Eu susțin că toate explicațiile sunt cauzale. A explica un fenomen (un *explanandum*) înseamnă a te referi la un fenomen anterior (*explanans*) care l-a cauzat. Atunci când pledez în favoarea explicației cauzale, nu intenționez să exclud posibilitatea explicației intenționale a comportamentului. Intențiile pot servi drept cauze. Un tip particular de explicație intențională este *explicația prin alegerea rațională*, pe care o voi discuta pe larg în capitolele următoare. Cu toate acestea, multe explicații intenționale se bazează pe presupoziția că agenții sunt, într-un fel sau altul, *irațional*i. În sine, iraționalitatea este doar o idee negativă sau reziduală, ea include tot ceea ce nu este rațional. Pentru ca această idee să aibă valoare explicativă trebuie să facem apel la forme specifice de iraționalitate, care au implicații specifice pentru comportament. În capitolul 12, de exemplu, voi enumera și ilustra unsprezece mecanisme care pot genera un comportament irațional.

Oamenii de știință explică uneori fenomenele prin apel la consecințe, și nu la cauze. De exemplu, aceștia pot susține că dușmăniile ancestrale își au explicația în faptul că mențin un nivel sustenabil al populației. Aceasta poate părea o imposibilitate metafizică: cum poate fi explicată apariția unui lucru la un anumit moment în timp prin ceva care încă nu există? După cum vom vedea, putem reformula problema astfel încât să facem din explicația prin consecințe un concept cu sens. Un exemplu este oferit de științele biologice: explicația evoluționistă. Totuși, cazurile în care o astfel de explicație are succes sunt foarte rare în științele sociale. Exemplul de mai sus, referitor la dușmăniile ancestrale, nu este cu siguranță unul dintre ele.

ROMANIAN ACADEMY UNIVERSITY LIBRARY

Științele naturii, în special fizica și chimia, ne oferă *explicația prin legi*; legile sunt propoziții generale care ne dau posibilitatea de a infera adevărul unui enunț relativ la un anumit moment din adevărul unui alt enunț relativ un moment anterior. Astfel, dacă știm pozițiile și viteza planetelor la un anumit moment, legile mișcării planetelor ne permit să deducem și să facem predicții cu privire la pozițiile lor la momente ulterioare. Acest tip de explicație are caracter *determinist*: dat fiind antecedentul, este posibilă o singură consecință. Științele sociale oferă foarte puține astfel de explicații prin legi (sau poate chiar niciuna). Relația dintre enunțul care explică (*explanans*) și lucrul care trebuie explicat (*explanandum*) nu este de tip „unu la unu" sau „mai mulți la unu", ci este de tip „unu la mai mulți" sau „mai mulți la mai mulți". În cadrul științelor sociale, se încearcă modelarea acestei relații prin metode *statistice*. Totuși, explicațiile statistice sunt incomplete în sine, în cele din urmă ele trebuind să recurgă la intuiții referitoare la *mecanisme* cauzale plauzibile.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 1

EXPLICAȚIA

Explicația în general

Sarcina fundamentală a științelor sociale este aceea de a explica fenomenele sociale. Nu este singura sarcină, dar este cea mai importantă, dat fiind că le subordonează pe celelalte, care depind de aceasta. Tipul fundamental de explanandum este un *eveniment*. A explica înseamnă a spune de ce a survenit acel eveniment prin referire la un *eveniment anterior*, considerat drept cauză a celui dintâi. Astfel, putem explica victoria lui Ronald Reagan în alegerile din 1980 prin încercarea nereușită a lui Jimmy Carter de a-i salva pe americanii luați ostatici în Iran.[1] Sau putem explica izbucnirea celui de-al Doilea Război Mondial referindu-ne la un număr de evenimente anterioare, de la acordul de la München până la semnarea Tratatului de la Versailles. Cu toate că, în ambele cazuri, structura de detaliu a explicației cauzale va fi, evident, mult mai complexă, acestea sunt instanțe ale modelului explicativ fundamental de tip *eveniment – eveniment*. În tradiția inaugurată de David Hume, acest model al explicației cauzale este supranumit „modelul bilelor de biliard". Un eveniment, bila A, lovește bila B, fiind cauza – și, astfel, explicația – unui alt eveniment, anume că bila B începe să se miște.

Cei care sunt familiarizați cu tipul curent de explicație în științele sociale nu vor recunoaște acest model sau nu îi vor recunoaște merite deosebite. Într-un fel sau altul, cercetătorii din științele sociale pun un accent mai puternic pe *fapte* sau stări de lucruri decât pe evenimente. Propoziția „La ora nouă dimineața drumul era alunecos" enunță un fapt. Propoziția „La ora nouă dimineața mașina a ieșit în decor" enunță un eveniment. După cum sugerează acest exemplu, se poate oferi o explicație de tip

---

[1] Pentru a anticipa o distincție pe care o voi discuta mai târziu, trebuie ținut cont de faptul că Jimmy Carter *nu a eșuat în încercarea sa*, ci *a încercat și a eșuat*. O non-acțiune, precum eșecul de a încerca, nu poate avea eficiență cauzală, cu excepția sensului indirect că, dacă alții percep sau inferează un eșec de a acționa, atunci pot îndeplini acțiuni pe care altfel nu le-ar fi îndeplinit.

ROMANIAN ACADEMY LIBRARY

fapt-eveniment pentru a da seamă de accident.[1] Putem inversa lucrurile propunând o explicație de tip *eveniment-fapt* pentru a da seamă de o anumită stare de lucruri, ca atunci când afirmăm că atacul de la World Trade Center din 2001 explică starea de frică generalizată în rândul americanilor. De asemenea, explicația-standard în științele sociale urmează modelul fapt-fapt. Un exemplu ales la întâmplare: s-a susținut că nivelul de educație al femeilor explică venitul *per capita* în țările în curs de dezvoltare.

Să luăm în considerare explicația unui fapt anume, acela că 65% dintre americani sunt de acord sau susțin că sunt de acord cu pedeapsa cu moartea.[2] În principiu, problema poate fi reformulată în termeni de evenimente: cum au ajuns acești americani *să fie de acord* cu pedeapsa cu moartea? Ce evenimente i-au format – interacțiunile cu părinții, colegii și profesorii – astfel încât să ajungă să aibă o astfel de atitudine? În practică, cei care lucrează în științele sociale nu sunt interesați de obicei de o astfel de întrebare. Ei nu caută neapărat o explicație care să dea seamă de o realitate statistică brută de acest fel, ci caută să înțeleagă *schimbarea* atitudinilor în timp sau *diferențele* de atitudine la nivelul mai multor populații. Motivul este, probabil, acela că nu consideră că faptul brut are o valoare informativă prea mare. Presupunând că ne-am întreba dacă 65% este mult sau puțin, cineva ar putea întreba „în comparație cu ce anume?". În comparație cu atitudinea americanilor în jurul anului 1990, când circa 80% erau de acord cu pedeapsa cu moartea, procentul actual este mic. În comparație cu atitudinea generală în anumite țări europene, este un procent mare.

Studiile *longitudinale* iau în considerare variații în timp ale variabilei dependente. Studiile *transversale* iau în considerare variații la nivelul populațiilor. În ambele cazuri, explanandumul este transformat, în sensul că nu se încearcă explicarea fenomenului „în sine", ci modul în care acesta variază în timp sau spațiu. Succesul unei explicații este măsurat, în parte, prin cât de mult poate da seamă de „varianță" (o unitate de măsură tehnică a variației).[3] Succesul complet ar explica toată variația observată. De exemplu, într-un studiu transversal efectuat pe întreaga națiune s-ar putea regăsi ideea că procentul acelora care sunt de acord cu pedeapsa cu moartea este strict proporțional cu numărul de omucideri la 100 000 de locuitori. Deși acest rezultat *nu* produce o explicație a numerelor absolute, ar oferi o explicație *perfectă* a diferenței dintre acestea.[4]

---

[1] Exemplul despre prezența la vot, discutat mai târziu, are în vedere același lucru.

[2] Răspunsurile fluctuează. De asemenea, numărul acelora care sunt de acord cu pedeapsa cu moartea pentru crimă scade drastic atunci când alternativa luată în considerare este închisoarea pe viață fără posibilitatea eliberării condiționate.

[3] După cum se exprimă economiștii, interesul este pentru ceea ce se întâmplă „în marjă".

[4] Riguros vorbind, lanțul cauzal ar putea să meargă în direcția opusă, de la atitudini la comportament, dar, în acest caz, acea ipoteză nu este plauzibilă.

În practică, firește, succesul total nu este atins niciodată, dar argumentul este valabil. Explicațiile varianței nu spun nimic despre explanandum „în sine".

Putem da un exemplu din studiul comportamentului la vot. După cum vom vedea mai târziu (capitolul 12), nu este clar de ce oamenii își exprimă votul în alegerile naționale când este aproape sigur că un singur vot nu poate să facă vreo diferență la nivel global. Cu toate acestea, o parte semnificativă a electoratului se prezintă la urne în ziua alegerilor. De ce?

În loc să se ocupe cu rezolvarea acestui mister, savanții din științele sociale empirice caută să răspundă la o întrebare diferită: de ce variază prezența la vot de la un scrutin la altul? O ipoteză ar fi aceea că votanții sunt mai puțin predispuși să iasă la urne dacă vremea nu este favorabilă, iar ploaia și frigul îi fac să stea acasă. Dacă datele se potrivesc cu ipoteza, după cum indică linia C din Figura 1.1, am putea susține că am explicat (cel puțin în parte) varianța prezenței la vot. Și totuși, nu vom fi oferit *nicio* explicație asupra faptului că linia C se intersectează cu axa verticală în punctul P mai degrabă decât în Q sau R. Este ca și cum am considera prima zecimală ca fiind dată și ne-am concentra pe explicarea celei de-a doua. Din punctul de vedere al prognozei, e posibil să nu fie nevoie de mai mult. Din punctul de vedere al explicației, suntem departe de a fi mulțumiți. „Evenimentul brut" că 45% sau mai mult din electorat vine de obicei la vot *este* interesant și reclamă insistent o explicație.

Procedura ideală, într-o perspectivă de tip eveniment-eveniment, ar fi următoarea: să luăm în considerare două tururi de scrutin, A și B. Pentru fiecare dintre acestea, avem de identificat evenimentele care cauzează un anumit procent al prezenței la vot. Odată ce am explicat astfel prezența la vot pentru scrutinul A, precum și prezența

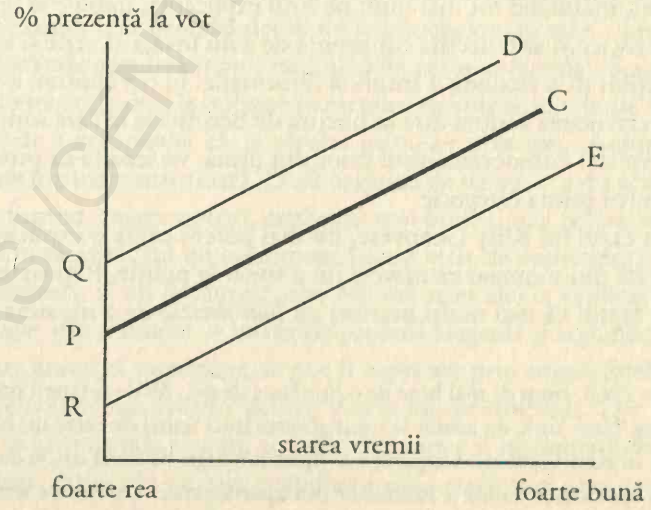

FIGURA 1.1

BCU IASI / CENTRAL UNIVERSITY LIBRARY

la vot pentru scrutinul B, explicația diferenței dintre cele două tururi de scrutin (dacă există vreuna) apare automat ca un produs secundar. Ca bonus, am fi în măsură să dăm o explicație a faptului dacă prezența la vot pentru cele două tururi de scrutin este accidentală, adică este cauzată de diferențe care se compensează reciproc sau nu. În practică, această procedură poate fi foarte greoaie. Datele sau teoriile disponibile s-ar putea să nu ne permită o explicație a fenomenelor „în sine". Trebuie să fim conștienți, însă, că, dacă recurgem la explicații ale variației, ne angajăm într-o practică explicativă de calitate secundară.

Uneori, cei care lucrează în câmpul științelor sociale încearcă să explice *non-evenimente*. De ce există atât de mulți oameni care nu își revendică beneficiile sociale la care sunt îndreptățiți? De ce nu a dat nimeni telefon la poliție în cazul Kitty Genovese?[1] Referitor la prima întrebare, o explicație poate fi că respectivii indivizi *decid* să nu revendice respectivele beneficii întrucât nu vor să fie stigmatizați sau pentru că sunt preocupați de imaginea de sine. Din moment ce o decizie *este* un eveniment, am avea de-a face, în acest caz, cu o dare de seamă satisfăcătoare. În cazul în care aceasta ar eșua, oamenii de știință ar investiga din nou *diferențele* dintre aceia care au dreptul la anumite beneficii și le revendică și aceia care au aceleași drepturi, dar nu le revendică. Să presupunem că singura diferență stă în faptul că aceștia din urmă nu sunt conștienți de existența vreunui drept. Ca explicație, acest aspect ne ajută, dar nu este suficient. Am vrea să mergem mai departe și să explicăm *de ce* anumiți indivizi nu sunt conștienți de faptul că au anumite drepturi. Ne-ar ajuta să știm, însă tot ar fi insuficient pentru o explicație dacă am afla că acești oameni sunt analfabeți și, prin urmare, incapabili să citească scrisorile care îi informează cu privire la drepturile lor. Însă, la un anumit moment, înaintând tot mai mult pe firul explicației, trebuie să ajungem la un eveniment pozitiv, cum ar fi decizia conștientă de a nu învăța să scrie și să citească sau decizia autorităților de a ascunde o anumită informație. În caz contrar, avem opțiunea de a recurge la cercetarea acelora care se bucură de beneficiile la care sunt îndreptățiți. Odată ce am explicat comportamentul celor din urmă, va rezulta ca produs secundar și explicația pentru prima categorie.

Referitor la cazul lui Kitty Genovese, nu mai putem căuta o explicație în variația comportamentală din moment ce *nimeni* nu a sunat la poliție. Rapoartele incidentului semnalează faptul că mai mulți martori *au luat decizia* de a nu suna la poliție. În

[1] În 27 martie 1964, timp de mai bine de o jumătate de oră, 38 de cetățeni pașnici și respectabili din Queens, New York, au asistat la înjunghierea unei femei de către un hărțuitor în trei atacuri separate, în Kew Gardens. Ucigașul s-a oprit și a fugit de două ori, o dată la auzul vocilor, apoi după aprinderea subită a luminilor din apartamente. De fiecare dată, însă, acesta s-a întors, și-a căutat victima și a înjunghiat-o din nou. Nimeni nu a sunat la poliție în timpul incidentului; unul dintre martori a dat telefon numai după ce femeia murise deja.

termeni de cauze proxime, aceasta este o dare de seamă satisfăcătoare, cu toate că am vrea să cunoaștem și motivele care au stat în spatele deciziei lor. Oare le-a fost frică „să se implice" sau fiecare s-a gândit că altcineva va suna la poliție, pe principiul „Copilul cu multe moașe rămâne cu buricul netăiat"? Totuși, se pare că unii martori nici măcar nu s-au gândit să sune la poliție. Un bărbat s-a uitat împreună cu soția lui la întregul episod doar pentru distracție, iar altcineva a declarat că era obosit și s-a dus la culcare. Pentru a explica de ce nu au avut o reacție mai puternică la ceea ce se întâmpla sub ochii lor, am putea aduce în discuție superficialitatea sentimentelor lor, însă aceasta ar însemna să recurgem la un explanans negativ pentru a da seamă de un explanandum la fel de negativ. Din nou, comportamentul acelor martori nu poate fi explicat altfel decât ca un produs derivat sau rezidual al explicației. Doar când vom avea o explicație satisfăcătoare a faptului că unii dintre ei s-au gândit să sune la poliție, chiar dacă în cele din urmă s-au decis să nu o facă, vom avea și o explicație a faptului că alții nici măcar nu s-au gândit să anunțe autoritățile.

Pe parcursul cărții, voi adopta în multe locuri o atitudine mai relaxată în raport cu această abordare puristă și rigoristă despre ce anume constituie un explanandum relevant și o explicație adecvată. Insistența pe explicații centrate pe eveniment seamănă puțin cu principiul individualismului metodologic, care este o altă premisă a cărții. În principiu, explicațiile în cadrul științelor sociale ar trebui să-i aibă în vedere doar pe indivizi și acțiunile lor. În practică, oamenii de știință utilizează destul de des entități supraindividuale, cum ar fi familiile, firmele sau națiunile, fie ca pe o *abreviere ajutătoare*, fie ca o *abordare secundară din punct de vedere calitativ*, la care sunt obligați să recurgă din cauza lipsei datelor relevante sau a unor teorii mai rafinate. Cele două justificări menționate se aplică și utilizării faptelor ca explananda sau explanantia, explicațiilor variației mai degrabă decât ale fenomenelor „în sine", precum și analizei acelor explananda negative (non-evenimentele sau non-faptele). Scopul discuțiilor precedente nu este acela de a le impune oamenilor de știință standarde imposibile sau fără noimă, ci de a argumenta că, la nivelul primelor principii, abordarea bazată pe evenimente este intrinsec superioară. Dacă oamenii de știință ar avea acest lucru în vedere, ar putea furniza, măcar uneori, explicații mai bune și mai fructuoase.

Uneori vrem să explicăm un eveniment (sau o serie de evenimente) recurgând la consecințele acestuia, și nu la cauzele sale. Nu mă refer aici la explicația prin consecințe *intenționate*, din moment ce intențiile precedă alegerile și acțiunile pe care le explică. Ideea este aceea că evenimentele pot fi explicate prin consecințele lor *concrete*, de obicei, acele consecințe *benefice* pentru cineva sau pentru ceva. Din moment ce o cauză trebuie să își preceadă efectul, această idee pare a fi incompatibilă cu explicația cauzală. Și totuși, explicația cauzală poate lua forma explicației prin consecințe, dacă există ceva de felul unei bucle de la consecințe înapoi la cauze. La început, un copil poate să plângă pentru că simte durere, dar, dacă plânsul atrage atenția părinților,

atunci copilul ar putea începe să plângă și mai tare decât ar fi fost cazul. În capitolele 16 și 17 voi susține că acest tip de explicație este, într-o anumită măsură, marginal cu privire la studiul comportamentului uman. În cea mai mare parte a cărții mă voi apleca asupra celei mai simple forme de explicație cauzală, cea în care explanansul – care poate include credințe și intenții orientate spre viitor – precedă apariția explanandumului.[1]

Pe lângă forma, pe deplin respectabilă, a explicației funcționale, care se bazează pe mecanisme de feedback specifice, există și forme de explicație cu o reputație mai puțin bună și care nu fac altceva decât să trimită spre producerea de consecințe benefice într-un anumit sens, pentru ca apoi să presupună, fără a argumenta de ce, că acestea sunt suficiente pentru explicarea comportamentului care le-a cauzat. Atunci când explanandumul este o *instanță*, cum ar fi o acțiune singulară sau un eveniment, acest tip de explicație eșuează din motive de natură strict metafizică. Să luăm un exemplu din biologie. Nu putem explica apariția unei mutații neutre sau dăunătoare observând că această apariție a reprezentat o condiție necesară pentru o mutație benefică ulterioară. Atunci când explanandumul este un *tip*, cum ar fi o caracteristică de comportament recurentă, această formă de explicație ar putea fi validă. Însă, dacă nu este susținută de un mecanism de feedback specific, ar trebui să o tratăm ca și cum nu ar fi validă. Antropologii au susținut, de exemplu, că răzbunarea este un comportament cu diverse consecințe benefice, de la controlul populației până la impunerea de norme într-un sistem descentralizat. (Pentru mai multe exemple vezi capitolul 22.) Presupunând că aceste beneficii chiar se produc, apariția acestora ar putea fi încă doar accidentală. Pentru a arăta că acestea apar într-un mod neaccidental, deci că susțin comportamentul răzbunător care le cauzează, trebuie neapărat să demonstrăm existența unui mecanism de feedback. Chiar și atunci când avem un astfel de mecanism, apariția inițială a explanandumului trebuie să fi fost generată de altceva.

Structura explicațiilor

În cele ce urmează, voi oferi o prezentare mai detaliată a explicației în științele sociale (și, cel puțin până la un punct, în știință). Primul pas poate fi neglijat foarte ușor: înainte să vrem să explicăm un fapt sau un eveniment trebuie să stabilim că faptul *este* propriu-zis un fapt și că evenimentul a avut loc. Montaigne spunea: „Am observat că, dacă le ceri oamenilor să dea seama de *fapte*, aceștia își petrec timpul încercând mai degrabă să găsească motive pentru acestea, și mai puțin încercând să vadă dacă

[1] În funcție de scop, trebuie să facem o distincție între explicații cauzale, intenționale și funcționale. Fizica utilizează doar explicații cauzale, biologia admite și explicații funcționale, iar științele sociale acceptă și explicații intenționale. Însă, la un nivel fundamental, toate explicațiile sunt cauzale.

sunt sau nu adevărate... Faptele sunt trecute cu vederea în favoarea preocupării pentru deducerea atentă a inferențelor. În mod normal, ei își pun problema astfel: «Cum se întâmplă acest lucru?» Dar oare chiar se întâmplă cutare lucru? Iată ce ar trebui să îi preocupe."

Astfel, înainte de a încerca să explicăm de ce, de pildă, rata sinuciderilor este mai mare într-o țară decât în alta, trebuie să ne asigurăm că, în cazul celei din urmă, nu avem de a face cu o micșorare deliberată a numărului de sinucideri, cauzată, poate, de motive religioase. Înainte de a ne propune să explicăm de ce Spania are o rată a șomajului mai mare decât Franța, trebuie să ne asigurăm că diferențele raportate nu sunt cauzate de definiții diferite ale șomajului sau de existența unei piețe negre foarte dezvoltate în Spania. Dacă vrem să explicăm de ce șomajul în rândul tinerilor este mai mare în Franța decât în Marea Britanie, trebuie să decidem dacă explanandumul îl constituie rata șomajului în rândul tinerilor care caută o slujbă sau rata șomajului în rândul tinerilor în general, incluzându-i aici și pe studenți. Dacă facem o comparație între șomajul din Europa și cel din Statele Unite, trebuie să ne decidem dacă explanandumul îl reprezintă șomajul în sens literal, care include și populația din închisori, sau șomajul în sens tehnic, care îi include doar pe aceia care își caută o slujbă.[1] Înainte să încercăm să explicăm de ce răzbunarea ia forma „ochi pentru ochi" („Îl ucid pe unul dintre ai tăi de fiecare dată când îl ucizi pe unul dintre ai mei"), trebuie să fim siguri că am observat corect lucrurile. Poate că răzbunarea ia, mai degrabă, forma „doi ochi pentru un ochi" („Ucid doi dintre ai tăi de fiecare dată când îl ucizi pe unul dintre ai mei"). Știința, în cea mai mare parte, caută să explice lucruri pe care le știm cu toții, lucru valabil și pentru științele sociale, însă avantajul este că știința are cum să demonstreze că unele dintre lucrurile pe care credem că le știm nu sunt adevărate. În acest caz, științele sociale pot căuta să explice *de ce* credem că știm lucruri care nu sunt adevărate, adăugând în felul acesta ceva la cunoașterea noastră în locul cunoașterii care s-a dovedit falsă.[2]

---

[1] În fiecare dintre cele două cazuri de mai sus, unii indivizi pot ajunge criminali sau studenți pentru că nu consideră că au vreo șansă să ocupe un loc de muncă, chiar dacă ar încerca. În funcție de scopul avut în vedere, aceștia ar putea fi incluși în rândul șomerilor, însă nu întotdeauna.

[2] Așa cum poate explica credințele populare reducându-le la non-fapte, știința poate, de asemenea, să explice modul în care credințele populare acceptă explicații false. De exemplu, cei mai mulți dintre bolnavii de artrită cred că durerea lor este provocată de vremea rea. Studiile arată, însă, că o asemenea conexiune nu există. Poate că ar trebui să încetăm să mai căutăm o legătură între vreme și durerea artritică și să ne concentrăm eforturile pentru a explica de ce suferinzii de artrită cred că această legătură există. Cel mai probabil, la un moment dat li s-a spus că ar exista o legătură, iar ulterior aceștia au observat cu mult mai mare atenție momentele în

Să presupunem că avem un explanandum admis pe scară largă, dar pentru care nu există încă o explicație definitivă – o *problemă paradoxală*. Această problemă poate fi un fapt surprinzător sau contraintuitiv ori poate doar o corelație neexplicată. Un exemplu la scară redusă poate fi următorul: „De ce din bibliotecile de la Oxford se fură mai multe cărți de teologie decât cărți pe orice alt subiect?" Un alt exemplu, tot la scară redusă, pe care îl voi cerceta imediat în detaliu, este „De ce, în zilele noastre, publicul aplaudă în picioare spectacolele de pe Broadway mai frecvent decât acum douăzeci de ani?"

Ideal ar fi ca aceste probleme să fie abordate urmând setul de cinci pași de mai jos. În practică, însă, pașii (1), (2) și (3) apar într-o ordine diferită. Ne putem juca cu mai multe ipoteze până când una dintre ele se va dovedi cea mai promițătoare, iar apoi putem căuta o teorie care să o justifice. Dacă implementăm pașii (4) și (5) așa cum trebuie, putem încă să avem un grad sporit de încredere în ipoteza preferată. Însă, din rațiuni pe care le voi discuta spre sfârșitul capitolului următor, ar fi indicat ca oamenii de știință să stabilească anumite limite libertății de a alege între diverse ipoteze.

1. Alege teoria – un set de propoziții cauzale legate unele de altele – care promite să ofere explicația cea mai bună.
2. Specifică ipoteza care aplică teoria la problemă, în sensul că explanandumul este o consecință logică a ipotezei.
3. Identifică sau concepe scenarii plauzibile care ar putea oferi explicații alternative, inclusiv în sensul că explanandumul este o consecință logică a fiecărui scenariu.
4. Respinge fiecare dintre aceste scenarii rivale recurgând la implicații suplimentare testabile care nu sunt observate în mod concret.
5. Întărește ipoteza propusă arătând că are implicații suplimentare testabile, de preferință ale unor „fapte noi", care sunt observate în mod concret.

Aceste proceduri definesc așa-numita *metodă ipotetico-deductivă*. Într-un caz anume, aceste proceduri pot lua forma prezentată în Figura 1.2. O voi ilustra recurgând la problema pusă de creșterea numărului de spectatori care aplaudă în picioare la spectacolele de pe Broadway. Nu mă bazez nici pe observații sistematice, nici pe experimente controlate, ci pe impresiile mele superficiale, confirmate de diverse articole de presă. Pentru scopurile de față, caracterul neriguros al explanandumului nu contează. Dacă, într-adevăr, pe Broadway aplauzele în picioare sunt mai des întâlnite acum decât în urmă cu douăzeci de ani, ce pași am urma pentru a explica acest lucru?

---

care credința lor este confirmată și cu mult mai puțină atenție pe acelea în care credința lor este infirmată.

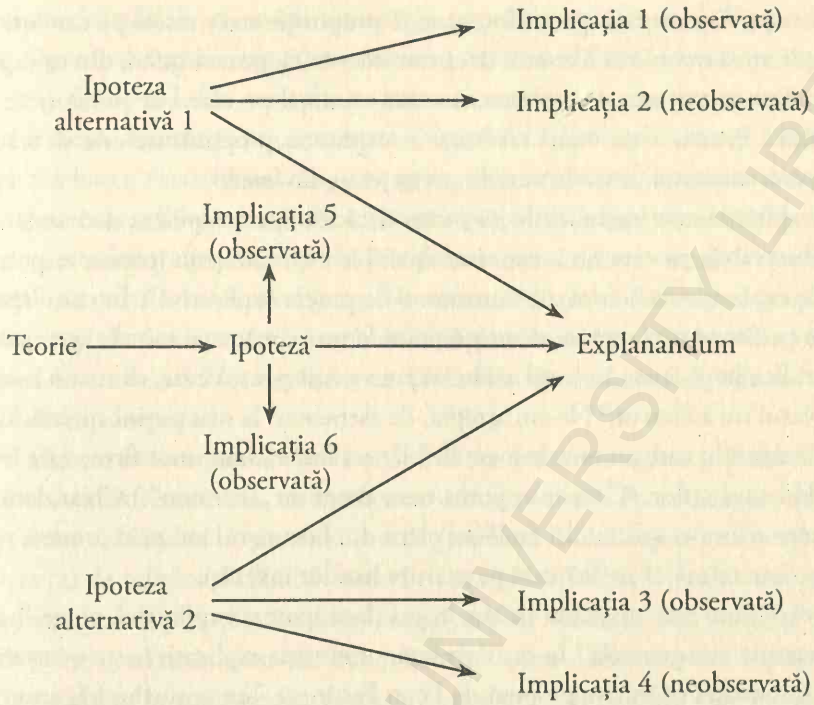

FIGURA 1.2

Voi încerca o explicație în funcție de prețul în creștere al biletelor pentru spectacolele de pe Broadway. Un cotidian relatează că dramaturgul Arthur Miller, ar fi spus: „Probabil că oamenii din public au impresia că, după ce au plătit 75 de dolari ca să stea jos, e rândul lor să se ridice în picioare. Nu vreau să par cinic, dar cred că totul s-a schimbat în momentul în care au crescut prețurile." Atunci când plătești 75 de dolari pentru un loc, e greu să admiți chiar și față de tine că ai urmărit un spectacol mediocru și, prin urmare, că ți-ai irosit banii. Aplauzele în picioare sunt o confirmare a faptului că spectacolul a fost o experiență plăcută.

Într-o prezentare mai formală, căutăm explicația în următoarea ipoteză: „Când oamenii investesc o sumă substanțială sau un efort substanțial pentru a obține un anumit bun, aceștia înclină (dacă totul merge cum trebuie) să prețuiască mai mult acel bun decât atunci când investesc o sumă mai mică pentru a-l procura."[1] Dată fiind

[1] O idee asemănătoare este utilizată uneori pentru a justifica tarifele mari practicate de psihoterapeuți: pacienții nu ar putea crede că terapia la care sunt supuși are succes decât dacă plătesc mult pentru o ședință. Dar, din câte știu, nu există vreun terapeut care să susțină că donează jumătate din tarif la Crucea Roșie.

BUCHAREST UNIVERSITY LIBRARY

premisa faptică a creșterii prețurilor, această propoziție trece testul pe care orice ipoteză explicativă trebuie să îl treacă: dacă este adevărată, putem infera din ea explanandumul. Dar acesta este, în realitate, un test minimal pe care l-ar putea trece multe propoziții.[1] Pentru a ne întări credința în explicația propusă, trebuie să arătăm că aceasta este susținută de jos în sus, de sus în jos și din lateral.

O explicație este susținută de jos în sus dacă din ipoteză putem deduce și verifica fapte observabile pe care nu aveam intenția să le explicăm prin ipoteza respectivă. O astfel de explicație trebuie să aibă un „exces de putere explicativă". În cazul spectacolelor de pe Broadway, trebuie să ne așteptăm la un număr mai mic de spectatori care se vor ridica în picioare în cazul acelor reprezentații pentru care, dintr-un motiv sau altul, prețul nu a crescut.[2] Ne-am aștepta, de asemenea, la mai puțini spectatori în picioare în cazul în care un număr mare de bilete a fost vândut unor firme care le-au făcut cadou angajaților. (Ceea ce ar putea trece drept un „fapt nou".) Chiar dacă aceste bilete sunt scumpe, spectatorii nu le-au plătit din buzunarul lor, prin urmare, nu sunt obligați să se mintă că se distrează pe măsura banilor investiți.

O explicație este susținută de sus în jos dacă ipoteza explicativă poate fi dedusă dintr-o teorie mai generală.[3] În cazul de față, propoziția explicativă este o instanțiere a teoriei disonanței cognitive propusă de Leon Festinger. Teoria susține că, atunci când o persoană resimte o inconsistență internă sau o disonanță la nivelul credințelor și valorilor, ne putem aștepta la o reajustare la nivel mental, care va elimina sau va reduce respectiva disonanță. În mod normal, ajustarea va alege calea minimei rezistențe. Cineva care a cheltuit 75 de dolari pentru un spectacol care se dovedește a fi mediocru nu se poate convinge foarte ușor că a cheltuit mai mult decât merită. Este mult mai simplu să se convingă de faptul că spectacolul a fost destul de reușit.

---

[1] Mintea omenească pare să aibă tendința de a transforma această condiție minimă într-o condiție suficientă. Odată ce ne-am fixat pe un scenariu care *ar putea* fi adevărat, de cele mai multe ori nu îl supunem niciunui test și nici nu luăm în considerare scenarii alternative. Alegerea unui scenariu se poate datora ideii de *post hoc ergo propter hoc* (se petrece după acest lucru, prin urmare se petrece din cauza acestui lucru) sau se poate datora unei inferențe care pornește de la faptul că un anumit scenariu este mai plauzibil decât altele, pentru a ajunge la concluzia că este *mai probabil* să fie corect.

[2] *Nu* ne așteptăm neapărat ca spectatorii din zonele cu bilete ieftine să stea jos. Aceștia s-ar putea simți prost să stea pe scaune când alții se ridică aplaudând frenetic. Ori s-ar putea ridica în picioare ca să-i vadă pe actori.

[3] Mai riguros spus: dacă este o instanțiere a unei teorii generale. Relația dintre o teorie generală și o anumită ipoteză explicativă este rareori una deductivă. În primul rând, teoria însăși ar putea avea anumite neajunsuri (vezi capitolul 2). În al doilea rând, de obicei o anumită teorie poate fi operaționalizată în moduri diferite.

Deși are dificultățile ei, teoria disonanței cognitive este destul de bine susținută empiric. O parte din sprijin vine dinspre cazuri destul de diferite de acela pe care l-am luat în considerare aici. De exemplu, o persoană care tocmai și-a cumpărat o mașină caută cu frenezie reclame pentru respectivul brand de mașină, pentru a-și întări convingerea că a luat o decizie bună. O parte din sprijinul acordat teoriei vine dinspre cazuri asemănătoare celui propus aici, ca, de exemplu, atunci când ritualurile de inițiere dureroase și umilitoare practicate în frățiile din campusuri duc la sentimente puternice de loialitate pentru respectivul grup. Nu vreau să sugerez că cineva și-ar spune în mod conștient: „Întrucât am suferit atât de mult pentru a adera la grup, acesta trebuie să fie unul foarte bun." Mecanismul prin care suferința determină loialitate este, cel mai probabil, inconștient.

O explicație este susținută din lateral dacă ne putem imagina și putem respinge ipoteze alternative care trec, de asemenea, testul minimal. Poate că avem de a face cu o sporire a aplauzelor în picioare întrucât publicul de astăzi este mai puțin sofisticat decât cel de odinioară, fiind format din spectatori din New Jersey, care ajung la teatru cu autobuzul, și nu din locuitori blazați ai New Yorkului. Sau poate că spectacolele de astăzi sunt realmente mai bune. Pentru fiecare dintre aceste alternative trebuie să ne gândim la anumite fapte care ar fi adevărate în cazul în care ipoteza ar fi adevărată, iar apoi să arătăm că aceste fapte nu sunt confirmate. Dacă publicul se ridică astăzi mai des în picioare ca să aplaude pentru că este mai ușor impresionabil, atunci ne-am aștepta ca, în trecut, publicul să se fi ridicat mai des în picioare la spectacolele desfășurate în afara orașului. Dacă spectacolele sunt mai bune decât odinioară, ne-am aștepta ca acest fapt să fie reflectat de cronicile de teatru și de numărul de reprezentații.

În cadrul acestei proceduri, apărătorul ipotezei originale trebuie să o facă și pe avocatul diavolului. Trebuie gândit în mod consistent *împotriva propriului sine* – să-ți faci viața cât mai grea cu putință. Ar trebui să alegem cea mai puternică și mai plauzibilă alternativă explicativă rivală, nu pe aceea ușor de respins. Din motive asemănătoare, atunci când căutăm să demonstrăm excesul de putere explicativă al unei ipoteze ar trebui să încercăm deducția și confirmarea unor implicații noi, contraintuitive și cât mai diferite de explanandumul original. Aceste două criterii – respingerea celei mai plauzibile alternative și generarea de *fapte noi* – sunt decisive pentru credibilitatea unei explicații. Sprijinul de sus în jos ne ajută, dar nu poate fi niciodată hotărâtor. Pe termen lung, teoria este susținută de explicațiile de succes pe care le generează, nu invers. Emilio Segrè, laureat al Premiului Nobel pentru Fizică, a spus că unii învingători conferă onoare premiului, în timp ce alții capătă onoare datorită premiului. Cei din urmă sunt paraziții propriului premiu. La fel, o teorie parazitează un număr de explicații

de succes pe care le generează. Este capabilă să confere sprijin unei anumite explicații doar pentru că a primit, la rândul ei, sprijin din partea unor explicații anterioare.

Ce nu constituie o explicație

Trebuie să distingem enunțurile care au sens explicativ de *alte șapte tipuri de enunț.*

În primul rând, trebuie să distingem explicațiile cauzale de *enunțurile cauzale adevărate.* A menționa o cauză nu este destul: trebuie furnizat (sau măcar sugerat) și mecanismul cauzal. În limbajul obișnuit, în romanele bune, în scrierile istorice de calitate și în multe analize din științele sociale mecanismul nu este prezentat în mod explicit. În schimb, acesta este sugerat de modul în care este descrisă cauza. Putem descrie în mai multe moduri orice eveniment dat. În explicațiile narative (de calitate) se presupune tacit că doar acele trăsături relevante cauzal ale unui eveniment sunt utilizate pentru a-l identifica. Dacă ni se spune că o persoană a murit deoarece a mâncat ceva stricat, presupunem faptul că mecanismul este acela al intoxicației alimentare. Dacă ni se spune că a murit în urma consumării de alimente la care era alergică, presupunem că mecanismul este constituit de o reacție alergică. Să presupunem acum că moartea a rezultat în urma unei intoxicații alimentare, dar că acea persoană era alergică la mâncarea respectivă, să zicem la homar. A spune că persoana a murit întrucât a mâncat ceva la care era alergică este adevărat, însă, în același timp, induce în eroare. A spune că a murit întrucât a consumat homar ar fi adevărat, dar neinformativ. Mecanismul cauzal nu ar fi nici măcar sugerat, cu toate că afirmația ar putea fi compatibilă cu mai multe astfel de mecanisme, ca, de exemplu, cu acela în care persoana este ucisă de cineva care a jurat să-l ucidă pe primul consumator de homari care îi iese în cale.

În al doilea rând, trebuie să distingem explicațiile cauzale de enunțurile referitoare la *corelații.* Uneori, suntem în situația în care putem spune că un eveniment de un anumit tip este invariabil sau urmat de obicei de un eveniment de alt tip. Acest lucru nu ne permite să spunem că evenimentele de primul tip sunt cauza acelora de al doilea tip întrucât există a treia posibilitate: cele două tipuri de evenimente pot fi efecte comune ale unui al treilea tip de eveniment. În *Viața lui Johnson*, Boswell descrie cum un anume Macaulay, cu toate că „avea o prejudecată împotriva prejudecăților", a afirmat că atunci când o navă sosește în St. Kilda, în Insulele Hebride, „toți locuitorii sunt cuprinși de răceală". Cu toate că s-au propus explicații cauzale ale acestui (așa-zis) fapt, un corespondent al lui Boswell l-a informat pe acesta că „localizarea St. Kilda face acostarea imposibilă fără existența unui vânt dinspre nord-est. Vântul, și nu străinul, este responsabil pentru epidemia de răceală." Putem lua în considerare și următoarea descoperire: copiii implicați în procese de custodie contestate la tribunal sunt mai labili psihic decât aceia ai căror părinți ajung la o înțelegere în mod privat. Am putea pune diferența chiar pe seama disputei asupra custodiei, întrucât le provoacă durere copiilor, precum și un sentiment de vină. Însă am putea pune diferența

și pe seama faptului că disputele asupra custodiei sunt purtate de părinți foarte ostili unul cu celălalt, iar copiii unor astfel de părinți tind să fie labili psihic. Pentru a putea distinge între cele două interpretări ar trebui să măsurăm nivelul suferinței înainte și după divorț. Voi descrie mai târziu și a treia posibilitate.

Să luăm în considerare un exemplu mai complex, de fapt, chiar exemplul meu favorit pentru acest tip de ambiguitate. În *Despre democrație în America*, Alexis de Tocqueville discută pretinsa legătură cauzală dintre căsătoria din dragoste și mariajul nefericit. El subliniază că această conexiune există doar în societățile în care acest tip de căsătorie este excepția, iar căsătoria aranjată este regula. Doar oamenii încăpățânați merg împotriva curentului general și este puțin probabil ca doi oameni încăpățânați să aibă o căsătorie fericită.[1] În plus, oamenii care merg împotriva curentului nu sunt atât de bine tratați de către membrii conformiști ai societății, iar acest lucru provoacă amărăciune și nefericire. Dintre argumentele prezentate, primul se bazează pe o corelație non-cauzală, datorată unui „factor terț", între căsătoria din dragoste și nefericire. Al doilea argument indică o conexiune cauzală adevărată, dar nu cea pe care o aveau în minte criticii căsătoriilor din dragoste, cărora li se adresa Tocqueville. Căsătoria din dragoste cauzează nefericire doar într-un context unde această practică este excepția. Biologii consideră adesea astfel de efecte ca fiind „dependente de frecvență".[2]

În plus față de problema „factorului terț", corelația ne lasă într-o stare de incertitudine cu privire la direcția cauzalității. Să luăm în considerare o glumă veche:

Psihologul: Ar trebui să te comporți frumos cu Johnny. Vine dintr-o familie dezorganizată.
Profesorul: Nu mă mir, Johnny ar putea dezorganiza orice familie.

Sau, cum spunea actorul de comedie Sam Levinson, „nebunia este ereditară. Poți s-o iei de la propriii copii." Sugestia este aceea că un copil labil psihic poate cauza

---

[1] În acest caz, „factorul terț" este o trăsătură de caracter, și nu un eveniment, respectiv încăpățânarea.

[2] Primul mecanism este un *efect de selecție*, al doilea este un *efect secundar* autentic. Această distincție are o mare aplicabilitate. Oricare dintre cele două mecanisme poate funcționa (poate chiar ambele) dacă ne întrebăm de ce o persoană aflată într-o anumită stare (având o anumită ocupație, fiind șomeră sau fiind spitalizată din cauza unei afecțiuni mentale) este cu atât mai probabil să rămână în acea stare cu cât nu a ieșit din ea de mai mult timp. Șomerii pe termen lung, de exemplu, pot forma un subset al populației cu abilități pentru care cererea este foarte redusă; există o probabilitate pentru orice angajat să-și piardă slujba, dar, după ce o pierde, starea de a fi șomer îl schimbă (sau percepția pe care o au angajatorii asupra lui), astfel încât probabilitatea ca el să reintre în câmpul muncii scade cu trecerea timpului. „Teoria etichetării" referitoare la bolile mintale sau la criminalitate se bazează pe asumpția (dubioasă) că efectele secundare domină efectele de selecție.

divorțul părinților, și nu că divorțul este cauza labilității psihice. În mod asemănător, o corelație negativă între cât de mult știu părinții despre ce fac copiii lor adolescenți și tendința acelor copii de a crea probleme nu indică neapărat că supravegherea copiilor de către părinți funcționează, ci doar că este puțin probabil ca adolescenții, al căror scop este acela de a crea probleme, să-și informeze părinții despre ceea ce fac.

În al treilea rând, trebuie să distingem explicația cauzală de enunțurile despre *necesitare*. A explica un eveniment înseamnă a răspunde la întrebarea de ce acel eveniment s-a produs *în modul în care* s-a produs. Ideea că acesta s-ar fi putut produce altfel și că s-ar fi produs altfel dacă nu s-ar fi produs, de fapt, în modul în care s-a produs nu ne oferă un răspuns la aceeași întrebare. Să luăm în considerare o persoană care suferă de cancer al pancreasului, care va determina cu siguranță moartea acelei persoane în decurs de un an. Când durerea devine insuportabilă, persoana se sinucide. Pentru a *explica* de ce aceasta a murit într-o anumită perioadă de timp nu are sens să spui că ea trebuia să moară în acea perioadă de timp deoarece avea cancer.[1] Dacă tot ce știm despre acest caz se rezumă la momentul apariției bolii, la durata limitată de viață pe care o au persoanele cu acest tip de cancer și la moartea persoanei, este plauzibil să inferăm că persoana a murit din cauza cancerului. Avem evenimentul anterior și mecanismul cauzal suficient pentru a produce evenimentul ulterior. Dar mecanismul nu este necesar: acesta poate fi înlocuit de altul. (În exemplul dat, cauza înlocuitoare este ea însăși un efect al cauzei înlocuite, dar lucrurile nu trebuie neapărat să stea astfel; persoana poate să moară și într-un accident de mașină.) Pentru a afla ce s-a întâmplat de fapt trebuie să cunoaștem mai multe detalii. Căutarea nu se termină niciodată: o cauză oarecare poate înlocui oricând cancerul.[2]

Enunțurile despre necesitare sunt uneori numite „explicații structurale". Analiza Revoluției Franceze întreprinsă de Tocqueville este un astfel de exemplu. În versiunea publicată a cărții, acesta enumeră o serie de evenimente și tendințe, începând cu secolul al XV-lea până în 1780, și afirmă că revoluția, având un astfel de fundal, era „inevitabilă". Prin aceasta a vrut probabil să spună (1) că orice serie de evenimente neînsemnate sau de importanță medie ar fi fost suficientă pentru a o declanșa și (2) că era aproape o certitudine că *unele* evenimente declanșatoare vor avea loc, deși nu neapărat acelor care au avut loc în realitate și nu neapărat în momentul în care au avut

---

[1] James Fitzjames Stephen scrie că „legea este foarte clară: dacă moartea cuiva se datorează unui atac survenit într-o fază timpurie a unei boli fatale care l-ar fi ucis într-un moment ulterior, atunci cauza morții este atacul".

[2] Trebuie să facem o distincție între înlocuirea unei cauze și supradeterminarea cauzală. Cea din urmă poate fi ilustrată de o persoană care este lovită simultan de două gloanțe: oricare dintre ele ar fi fost suficient pentru a o ucide. Cea dintâi este ilustrată de o persoană ucisă de un glonț, devansându-l pe altul, tras câteva secunde mai târziu.

loc. Autorul pare să susțină, de asemenea, că (3) după 1750, sau poate 1770, nu se mai putea face nimic pentru a împiedica izbucnirea revoluției. Cu toate că Tocqueville a lăsat în urmă notițe pentru un al doilea volum, în care intenționa să descrie modul în care s-a produs de fapt revoluția, s-ar putea argumenta că, dacă acesta a demonstrat punctele (1), (2) și (3), acest pas suplimentar nu mai era necesar. Problema cu acest raționament este că, în cazul multor întrebări interesante din științele sociale (și prin contrast cu exemplul cu cancerul), aserțiuni de tipul (1), (2) sau (3) sunt foarte greu de demonstrat prin metode care nu sunt alterate de o percepție ulterioară asupra evenimentelor.[1] Se poate da un argument mai puternic în cazul unor evenimente care se întâmplă independent unul de altul, dar, în același timp, sugerând că acestea „pluteau cumva în aer". Studiul unor descoperiri simultane în știință ne oferă un astfel de exemplu.

În al patrulea rând, trebuie să distingem explicația cauzală de *istorisire*. O explicație autentică dă seamă de ceea ce s-a întâmplat în felul în care s-a întâmplat. A spune o poveste înseamnă a da seamă de ceva în felul în care *s-ar fi putut* întâmpla (și poate chiar s-a întâmplat). Tocmai am argumentat că explicația științifică diferă de descrieri ale evenimentelor care *trebuiau* să se întâmple. Acum susțin că aceasta este diferită și de descrierile modului în care lucrurile *s-ar fi putut* întâmpla. Argumentul poate părea unul banal sau ciudat. De ce ar vrea cineva să propună o dare de seamă pur conjecturală a unui eveniment? Oare își au locul în știință astfel de speculații? Răspunsul este da, însă locul lor nu trebuie confundat cu acela al explicațiilor.

Istorisirile pot sugera explicații noi și parcimonioase. Să presupunem că cineva afirmă că sacrificiul de sine sau comportamentul altruist constituie o dovadă concludentă că nu toate acțiunile sunt centrate pe propriul interes și că existența comportamentului irațional constituie o dovadă concludentă că nu toate acțiunile sunt raționale. Am putea trage de aici concluzia că există trei forme ireductibile de comportament: rațional și egoist, rațional și altruist și irațional. Însă parcimonia care caracterizează demersurile științifice serioase ar trebui să ne facă suspicioși față de o asemenea perspectivă. Nu este oare posibil ca atunci când oamenii își ajută semenii să o facă așteptând reciprocitate, iar atunci când se enervează să o facă pentru a-și impune un anumit punct de vedere? Printr-o istorisire despre modul în care interesul personal rațional *ar putea* genera comportamente altruiste și emoționale reușim transformarea

---

[1] Revoluția Americană este, poate, un candidat mai plauzibil la o explicație structurală. Ministrul francez Choiseul, un observator neutru, dar cu un simț acut de observație, a remarcat încă din 1765 că independența coloniilor americane era inevitabilă. Un comentator francez detașat precum Raymond Aron putea conchide că Algeria își va câștiga independența cu mult înainte ca acest lucru să se producă. Revoluția Franceză seamănă mai degrabă cu căderea comunismului – inevitabilă, mai ales privită retrospectiv.

unei chestiuni filosofice într-una care poate fi călăuzită de cercetarea empirică.[1] O istorisire (neverificabilă și nefalsificabilă) poate fi primul pas în construcția unei explicații de succes. De fapt, multe dintre „răspunsurile" pe care le voi oferi în concluzii la probleme pe care le-am prezentat în introducere au caracterul unor asemenea istorisiri.

În același timp, istorisirea poate induce în eroare și poate face mult rău atunci când este confundată cu explicația autentică. Cu două excepții, pe care le voi enunța în paragraful următor, explicațiile de tipul „ca și cum" nu explică, de fapt, nimic. Să luăm în considerare o teză des întâlnită, și anume aceea că putem utiliza modelul alegerii raționale pentru a explica comportamentul, chiar dacă *știm* că oamenii nu pot efectua calculele mentale complexe pe care le presupune modelul (și pe care le regăsim în anexele din revistele de specialitate, unde este prezentat respectivul model). Cât timp modelul face predicții care se potrivesc destul de bine cu comportamentul observat, suntem îndreptățiți (se susține) să presupunem că agenții se comportă „ca și cum" ar fi raționali. Aceasta este viziunea operaționalistă sau instrumentalistă a explicației, care își are originea în fizică și care a fost adoptată ulterior de Milton Friedman în științele sociale. Se susține că motivul pentru care putem presupune că un bun jucător de biliard cunoaște regulile fizicii și poate efectua calcule complexe în minte este acela că această presupunere ne permite să-i prezicem și să-i explicăm comportamentul cu o acuratețe foarte mare. A întreba dacă presupunerea în sine este *adevărată* înseamnă a rata miza chestiunii.

Acest argument poate fi valid în anumite situații, în care agenții pot învăța prin încercare și eroare în timp. Dar este valid tocmai pentru că putem identifica un *mecanism* care determină în mod neintenționat același rezultat pe care un agent supraraional l-ar fi calculat intenționat.[2] În absența unui astfel de mecanism, am putea accepta încă perspectiva instrumentalistă în cazul în care presupoziția ne-ar permite să prezicem comportamentul cu mare acuratețe. Legea gravitației a fost un mister o perioadă lungă de timp, întrucât părea să se bazeze pe ideea neinteligibilă de acțiune la distanță. Dar, întrucât permitea predicții precise până la câteva zecimale, teoria lui Newton a fost acceptată unanim până la teoria relativității generale. Modul misterios

---

[1] În acest caz particular se întâmplă ca istorisirile să fie false din moment ce oamenii își pot ajuta semenii cu care nu vor mai interacționa niciodată, iar mânia unei persoane poate cauza reticență în rândul semenilor cu care ar putea interacționa.

[2] Trebuie să precizez în acest loc, unde utilizez pentru prima dată în carte cuvântul „agent", că mulți oameni de știință preferă termenul „actor". Poate economiștii gândesc în termeni de agenți, iar sociologii în termeni de actori. Cu toate că nu are o importanță prea mare ce termen folosim, eu prefer „agent" întrucât sugerează ideea de acțiune; prin contrast, „actor" sugerează un public care poate sau nu să fie de față.

în care funcționează mecanica cuantică este, de asemenea, acceptat, în ciuda unei anumite nesiguranțe, întrucât ne dă posibilitatea să facem predicții de o acuratețe chiar mai uimitoare decât modelul lui Newton.

Științele sociale care utilizează modelul alegerii raționale nu se pot baza pe un astfel de sprijin. *Nu există niciun mecanism general neintenționat care poate simula sau imita raționalitatea.* Învățarea prin consolidare (capitolul 16) poate face acest lucru în unele cazuri, cu toate că în alte cazuri produce devieri sistematice de la raționalitate. Un soi de analogon social al selecției naturale ar putea să o facă în alte cazuri, în linii mari, cel puțin, dacă viteza de schimbare a mediului este mai mică decât viteza de ajustare (capitolul 17). În situațiile irepetabile sau în medii care se schimbă rapid, nu cunosc vreun mecanism care ar putea să simuleze raționalitatea. În același timp, sprijinul empiric pentru explicațiile prin alegerea rațională ale unor fenomene complexe tinde să fie relativ redus. Firește, aceasta este o afirmație generală. Nu voi explica ce înțeleg prin „redus", vreau doar să scot în evidență gradul mare de dezacord în care se găsesc oameni de știință competenți cu privire la forța explicativă a unor ipoteze concurente. Chiar și în economie, oarecum cea mai dezvoltată dintre științele sociale, există dezacorduri fundamentale și persistente între diverse „școli de gândire". *Niciodată* nu observăm nivelul de precizie (la zecimală) care ar pune capăt controverselor.

În al cincilea rând, trebuie să distingem explicația cauzală de *explicația statistică.* Cu toate că multe explicații în științele sociale sunt de ultimul tip, acestea sunt nesatisfăcătoare întrucât nu pot da seamă de evenimente individuale. A aplica generalizări statistice la cazuri individuale este o gravă eroare, nu doar în știință, ci și în viața de zi cu zi.[1] Să presupunem că este adevărat că bărbații au tendința de a fi mai agresivi decât femeile. Să-i spui unui bărbat mânios că furia îi este produsă de hormonii masculini, în loc să-i spui că este nejustificată în momentul respectiv, înseamnă să comiți atât o eroare intelectuală, cât și una morală. Este o eroare intelectuală întrucât presupunem că o generalizare validă pentru cele mai multe cazuri este validă pentru fiecare caz în parte; este o eroare morală întrucât îl tratăm pe interlocutor ca pe un individ guvernat de mecanisme biologice, care nu este deschis la dialog și nu poate înțelege un argument.

---

[1] Eroarea opusă – utilizarea unui caz individual în sprijinul unei generalizări – trebuie evitată, de asemenea. Proust descrie modul în care menajera Françoise din familia naratorului „era la fel de probabil să considere cazul particular ca fiind general, precum era probabil să considere cazul general ca fiind particular". O astfel de combinație poate fi foarte dăunătoare. Să presupunem că observi un membru al grupului X spunând o minciună. Generalizând, îți formezi opinia că membrii grupului X tind să spună minciuni. Apoi, observând un alt membru al grupului, presupui că acesta minte. În cele din urmă, presupunerea (neverificată) este folosită ca probă pentru generalizare.

Cu toate că explicațiile statistice nu sunt cel mai bun tip de explicație, în practică este posibil să nu putem produce mai mult. Totuși, trebuie observat că și aceste explicații sunt ghidate de idealul explicației cauzale. Pare a fi un fapt statistic că, într-o democrație, cetățenii trăiesc mai mult decât cei din regimuri nedemocratice. Înainte de a trage concluzia că regimul politic explică longevitatea, poate ar fi bine să *controlăm* dacă nu cumva alte variabile sunt răspunzătoare pentru rezultat. Este posibil ca proprietatea X să fie prezentă mai mult în rândul democrațiilor decât în rândul regimurilor non-democratice și este posibil ca aceasta să fie responsabilă pentru speranța de viață. Însă, întrucât există un număr indefinit de mare de astfel de proprietăți, cum putem ști pe care anume să o controlăm? Răspunsul evident este acela că trebuie să ne ghidăm după o ipoteză cauzală. De exemplu, este cel puțin plauzibil că în societățile industrializate cetățenii trăiesc mai mult decât cei din societățile mai puțin dezvoltate. Dacă societățile industrializate au, de asemenea, tendința să fie mai democratice decât regimurile neindustrializate, acest lucru ar putea da seamă de faptele observate. Pentru a ne asigura că factorul cauzal este democrația, și nu industrializarea, s-ar putea să fie nevoie să comparăm regimuri democratice cu unele nedemocratice situate la același nivel de industrializare și să urmărim dacă înregistrăm vreo diferență. Odată ce suntem suficient de încrezători că am adus sub control alte cauze plauzibile, putem încerca să descoperim *cum* – prin ce mecanism sau lanț cauzal – afectează tipul de regim durata vieții. Voi discuta acest al doilea pas în capitolul următor. Pentru moment, vreau doar să observ că gradul nostru de încredere se bazează inevitabil pe *intuiții cauzale* privitoare la modul în care ținem anumite variabile sub control pentru a identifica „factorul terț" care este cel mai plauzibil.[1]

În al șaselea rând, trebuie să distingem explicațiile de *răspunsurile* la întrebarea „de ce?". Să presupunem că citim un articol de specialitate și observăm surprinși că autorul nu citează un alt articol foarte important și relevant, ceea ce ne face să întrebăm: „De ce nu citează acest articol?" Curiozitatea noastră poate fi satisfăcută pe deplin dacă aflăm că autorul nu cunoștea articolul respectiv (cu toate că am vrea să știm și de ce nu a făcut o cercetarea mai temeinică a literaturii de specialitate). Dar a spune „Nu a citat articolul întrucât nu îl cunoștea" nu este o explicație. Dacă interpretăm acest lucru ca pe o explicație, ar însemna să ne referim la un non-eveniment pentru a explica un non-eveniment, ceea ce este absurd. Să presupunem, însă, că descoperim că autorul cunoștea, de fapt, articolul, dar *a decis* să nu îl citeze întrucât nici el nu fusese

[1] De exemplu, nu există niciun mecanism cauzal care ar trebui să ne facă să supunem controlului dimensiunea populațiilor în regimurile democratice și nedemocratice. Cu toate că nu putem exclude de la bun început o legătură cauzală între dimensiunea populației și durata vieții, o astfel de legătură nu s-a făcut în nicio știință socială și nici nu îmi pot imagina o astfel de legătură care să nu fie pe de-a-ntregul născocită.

citat în cadrul articolului respectiv. În acest caz, răspunsul la întrebarea „de ce?" oferă în același timp și o explicație. Există un eveniment, decizia de a nu cita articolul, cauzat de un eveniment anterior, supărarea provocată de faptul că nu a fost citat.

În cele din urmă, trebuie să distingem explicația cauzală de *predicție*. Uneori putem explica fără să fim în stare să formulăm o predicție, alteori putem formula predicții fără să putem explica. Ce-i drept, în multe cazuri, una și aceeași teorie ne va permite să facem ambele lucruri, dar cred că în științele sociale aceasta este de cele mai multe ori excepția, nu regula.

Voi discuta în capitolul următor de ce putem avea o putere explicativă considerabilă fără o putere predictivă pe măsură. Pe scurt, motivul este acela că în multe cazuri putem identifica un mecanism cauzal după ce s-a produs un fapt, dar nu putem prezice mecanismul care se va declanșa (dintr-o mulțime de mecanisme plauzibile) înainte ca faptul să se întâmple. Cazul particular al explicației în biologie este întrucâtva diferit. După cum voi discuta în capitolul 16, evoluția este alimentată de mecanismele îngemănate ale mutației aleatorii și ale selecției (mai mult sau mai puțin) deterministe. Dată fiind o anumită trăsătură sau caracter comportamental al unui organism, îi putem explica *originea* recurgând la o schimbare aleatoare în materialul genetic și *persistența* prin impactul favorabil asupra capacității de adaptare. Cu toate acestea, anterior apariției mutației, nimeni nu ar fi putut să o prezică. Mai mult, întrucât apariția unei mutații determină mutațiile apărute ulterior, nu putem nici măcar să prezicem că o anumită mutație se va produce mai devreme sau mai târziu. Prin urmare, este puțin probabil ca explicațiile structurale să aibă succes în biologie. Fenomenul *convergenței* – specii diferite dezvoltă elemente de adaptare similare pentru că sunt supuse acelorași presiuni din partea mediului – pare a avea un caracter structural, însă nu ne permite să spunem că elementele de adaptare au apărut în mod inevitabil.

Invers, putem avea putere predictivă fără să avem și putere explicativă. Putem prezice un consum mai redus al unui bun dacă prețul acestuia crește, fără a fi nevoiți să formulăm o ipoteză pentru a explica un astfel de comportament al consumatorilor. Oricare ar fi originea acțiunii la nivel individual – rațională, tradițională sau, pur și simplu, aleatoare –, putem prezice că, în general, oamenii vor cumpăra mai puțin din bunul respectiv pentru simplul fapt că nu și-l mai permit (capitolul 9). Aici sunt implicate mai multe mecanisme pe care le constrângem să ducă la același rezultat, astfel încât, pentru scopuri care țin de predicție, nu trebuie să facem o alegere între ele. Cu toate acestea, pentru scopuri care țin de explicație contează mecanismul întrucât oferă înțelegere, în timp ce predicția oferă, în cel mai bun caz, doar control.

De asemenea, pentru scopuri care țin de predicție, distincția dintre corelație, necesitate și explicație este inutilă. Dacă există o regularitate legică între un anumit tip de evenimente și alt tip, atunci este irelevant – pentru scopuri care țin de predicție – dacă

faptul se datorează unei relații cauzale între aceste tipuri sau existenței unui al treilea tip, care le este cauză comună. În fiecare caz putem folosi apariția primului eveniment pentru a prezice apariția celui de-al doilea eveniment. Cu toate că nimeni nu crede că primele simptome ale unei boli sunt cele care cauzează ulterior moartea, acestea sunt folosite pentru a prezice acest eveniment. În mod asemănător, cunoscând starea de sănătate a unei persoane, putem prezice moartea acesteia în cel mult un an, iar această predicție nu este falsificată în cazul în care persoana se întâmplă să moară mai devreme într-un accident de mașină sau dacă se sinucide pentru că suferința provocată de boală devine de neîndurat.

◈◈◈

Notă bibliografică

Viziunea generală privitoare la explicație și cauzalitate pe care mă bazez este descrisă în detaliu în J. Elster, D. Føllesdal și L. Walløe, *Rationale Argumentation* (Berlin, Gruyter, 1988). Pentru aplicații asupra acțiunii umane, cititorul poate face apel la D. Davidson, *Essays on Actions and Events* (Oxford University Press, 1980). Critica pe care o fac explicației funcționale este prezentată în mai multe locuri, dintre care aș semnala *Explaining Technical Change* (Cambridge University Press, 1983). Detalii despre cazul lui Kitty Genovese se găsesc în A.M. Rosenthal, *Thirty-Eight Witnesses* (Berkeley, University of California Press, 1999). O prezentare accesibilă a teoriei lui Festinger se găsește în L. Festinger, S. Schachter și M. Gazzaniga (editori), *Extending Psychological Frontiers: Selected Works of Leon Festinger* (New York, Russell Sage, 1989). Exemplele legate de efectul copiilor asupra părinților le-am luat din două cărți foarte interesante de J.R. Harris, *The Nurture Assumption: Why Children Turn Out the Way They Do* (New York, Free Press, 1998) și *No Two Alike* (New York, Norton, 2006). Discut teoria lui Tocqueville despre cauzalitate în „Patterns of causal analysis in Tocqueville's *Democracy in America*", *Rationality and Society*, 3 (1991), pp. 277-297, și viziunea lui despre Revoluția Franceză în „Tocqueville on 1789", în C. Welch (ed.), *The Cambridge Companion to Tocqueville* (Cambridge University Press, 2006). Apărarea pe care o face Milton Friedman raționalității de tipul „ca și cum" în „The methodology of positive economics" (1953) este retipărită în M. Brodbeck (ed.), *Readings in the Philosophy of the Social Sciences* (Londra, Macmillan, 1969). O apărare recentă a abordării „ca și cum" în științele politice se regăsește în R. Morton, *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science* (Cambridge University Press, 1999). La fel ca marea majoritate a acelora care apără această abordare, autoarea nu oferă un motiv pentru care ar trebui să *credem* ficțiunea

„ca și cum". Întâlnim o excepție de la această regulă, însă numai parțial, la D. Satz și J. Ferejohn, „Rational Choice and Social Theory", *Journal of Philosophy*, 91 (1994), pp. 71-87. Discuția referitoare la întrebările de tipul „de ce?" se bazează pe B. Hansson, „Why explanations", în *Theoria*, vol. 72, nr. 1 (2006), pp. 23-59. Independența legii cererii de presupozițiile motivaționale a fost observată în G. Becker, „Irrational behavior in economic theory", *Journal of Political Economy*, 70 (1962), pp. 1-13.

BCU IASI CENTRAL UNIVERSITY LIBRARY

Capitolul 2

MECANISMELE

Deschiderea cutiei negre

Filosofii științei afirmă adeseori că o explicație trebuie să se bazeze pe o *lege cu caracter general*. A explica un eveniment înseamnă a menționa un set de condiții inițiale, împreună cu un enunț al cărui sens este acela că, ori de câte ori condițiile respective sunt satisfăcute, urmează să se întâmple un eveniment de acest tip. În acest capitol voi prezenta două obiecții la această idee, una moderată și relativ necontroversată, alta mai radicală și mai discutabilă.

Prima obiecție este aceea că, în cazul în care am stabilit o lege cu caracter general din care putem deduce explanandumul (a doua obiecție neagă faptul că putem face întotdeauna aceasta), acest lucru nu echivalează cu o explicație. Încă o dată: ne putem referi la distincția dintre explicație, pe de o parte, și corelație și necesitare, pe de alta. O lege cu caracter general care spune că anumite simptome ale unei boli sunt urmate invariabil de moarte nu explică de ce a murit o anumită persoană. O lege cu caracter general care se bazează pe natura fundamentală a bolii nu explică moartea dacă boala a fost înlocuită, la nivel cauzal, de sinucidere sau de un accident de mașină.

Pentru a evita aceste probleme, se susține adeseori că ar trebui să înlocuim ideea de lege cu caracter general cu ideea de *mecanism*. De fapt, întrucât voi folosi mai încolo termenul de „mecanism" într-un sens specializat, voi utiliza expresia „lanț cauzal" pentru a arăta ceea ce am în minte aici.[1] În loc să încercăm să explicăm evenimentul E prin enunțul „Ori de câte ori au loc evenimentele C_1, C_2,..., C_n, urmează un eveniment de tip E", putem stabili lanțul cauzal care pleacă de la cauzele C_1, C_2, ..., C_n până la E. Acest pas poartă de multe ori denumirea de „deschiderea cutiei negre". Să presupunem că știm faptul că este mult mai probabil ca fumătorii înrăiți să

[1] În unele lucrări timpurii am folosit termenul „mecanism" pentru a denumi ceea ce acum numesc „lanț cauzal". În lucrări mai recente am început să folosesc termenul „mecanism" în sensul pe care îl voi defini în acest capitol. Probabil că ar fi trebuit să aleg o terminologie diferită, dar acum este prea târziu.

BCU Cluj / UNIVERSITY LIBRARY

se îmbolnăvească de cancer pulmonar. Aceasta se poate datora fie faptului că fumatul este o cauză a cancerului la plămâni, fie faptului că oamenii predispuși la fumat sunt, de asemenea, predispuși la cancer la plămâni (poate că genele care predispun la cancer pulmonar sunt legate de genele care îi fac pe oameni dependenți de nicotină).[1] Pentru a stabili prima explicație ar trebui să punem în evidență un lanț de relații fiziologice de tip cauză-efect care începe cu fumatul masiv și se sfârșește cu cancerul la plămâni. Explicația finală va fi mai rafinată, va avea mai multe conexiuni cauzale și va fi mult mai convingătoare decât enunțul de tip cutie neagră „Fumatul cauzează cancer".

Sau să presupunem că cineva afirmă că o rată mare a șomajului cauzează războaie de agresiune și aduce dovezi care susțin o conexiune legică între cele două fenomene. Din nou: cum putem stabili că avem de-a face cu un efect cauzal, și nu cu o simplă corelație? Poate că rata mare a fertilității care cauzează șomajul îi motivează pe liderii politici să pornească războaie de agresiune. Războaiele soldate cu înfrângeri ar determina măcar reducerea populației, iar victoria ar oferi noi teritorii pentru expansiune și migrație. Pentru a elimina această posibilitate, ar trebui mai întâi să controlăm variabila reprezentată de rata fertilității (și alți „factori terți" plauzibili) și să vedem dacă se menține conexiunea. Dacă se menține, tot nu vom fi mulțumiți până nu ni se va permite accesul în cutia neagră și nu ni se va descrie *modul* în care rata mare a șomajului cauzează războaie. Aceasta se întâmplă deoarece șomajul îi face pe liderii politici să caute noi piețe de desfacere în urma războiului? Sau liderii politici cred că șomajul creează tensiuni sociale care trebuie canalizate spre un dușman extern pentru a împiedica mișcări revoluționare la nivel intern? Ori își imaginează că industria de armament poate absorbi o parte din șomaj? Sau poate că șomerii au tendința să voteze lideri populiști care, foarte probabil, vor ocoli mijloacele diplomatice și vor folosi războiul pentru rezolvarea conflictelor.

Să luăm în considerare mai în detaliu ultima propunere. De ce ar vota șomerii lideri populiști iresponsabili în locul unor politicieni din partide politice conservatoare? Din nou, ne putem imagina mai multe moduri prin care putem deschide această cutie neagră. Probabil că votanții obișnuiți ai politicienilor populiști se mobilizează la urne atunci când nu au un loc de muncă întrucât costul de oportunitate al votului (adică valoarea timpului petrecut la vot) este mai mică decât dacă ar avea un loc de muncă. Sau poate că liderii populiști vor propune soluții imediate pentru rezolvarea problemei șomajului. Sau poate că aceștia vor propune strategii care să îi pedepsească pe aceia pe care șomerii îi consideră responsabili pentru condiția lor (capitaliștii sau vreo minoritate etnică prosperă din punct de vedere financiar).

---

[1] După cum am observat mai devreme, a doua explicație a fost propusă cu toată seriozitatea la un anumit moment.

Să analizăm mai în detaliu ultima propunere. *De ce* ar vrea șomerii să îi pedepsească pe capitaliști sau vreo minoritate avută? Oare nu este și acesta un enunț de tipul cutiei negre? Putem dezvolta această teză completând-o cu afirmația că șomerii sunt motivați de interesul personal material. Dacă statul ar putea să confiște averea acestei elite, fondurile ar putea fi utilizate pentru a-i retribui pe șomeri. Sau poate că aceștia sunt motivați de dorința de răzbunare, caz în care ar putea fi incitați la pedepsirea respectivei elite chiar dacă nu ar avea beneficii materiale. Dacă există percepția că bogații recurg la reduceri de personal pentru a-și crește profiturile, atunci aceia care își pierd locurile de muncă se pot răzbuna la urne. Sau poate că șomerii sunt pur și simplu invidioși pe membrii inteligenți ai minorității care s-au bucurat de succes, în timp ce ei au eșuat, și se folosesc de vot pentru a-i pune la punct.

Din câte știu, o rată mare a șomajului nu cauzează războaie de agresiune. Întregul exercițiu este pur ipotetic, însă cred că vine în sprijinul ideii că credibilitatea unei explicații crește pe măsură ce legile generale sunt dezvoltate în termeni de lanțuri cauzale. La nivelul legilor generale, nu putem fi niciodată siguri dacă am controlat toți „factorii terți" relevanți. Există întotdeauna posibilitatea ca o cauză ascunsă să dea seamă atât de explanandum, cât și de pretinsa cauză. Dacă sporim numărul de conexiuni ale lanțului cauzal, atunci reducem acest pericol.

Cu toate acestea, pericolul nu poate fi înlăturat. Identificarea unui lanț cauzal nu presupune renunțarea completă la legile generale, ci doar coborârea de la legi generale cu caracter foarte abstract la legi cu un nivel de abstractizare mai mic. De exemplu, am putea înlocui legea universală „Rata mare a șomajului cauzează războaie" cu legea mai puțin abstractă „Liderii populiști sunt înclinați să poarte război" și „Șomerii votează lideri populiști". Cea din urmă lege, la rândul ei, poate fi înlocuită de o conjuncție dintre „Șomerii sunt invidioși pe minoritățile bogate" și „Cei care invidiază minoritățile bogate votează lideri populiști". La fel cum se poate întâmpla cu orice lege, și acestea se pot dovedi a fi simple corelații. Dacă a fi invidios pe minorități și a fi șomer ar fi efecte ale unei cauze comune, atunci succesul electoral al liderilor înclinați să poarte război nu s-ar datora șomajului, ci unui alt factor, corelat cu șomajul. Totuși, la acest nivel, mai rafinat, al demersului există mai puține variabile pe care le putem controla. Cu cât vom scoate mai bine în evidență istorisirea cauzală, cu atât ne va fi mai ușor să ne asigurăm că nu avem de-a face cu o simplă corelație.

Explicațiile în termeni de legi (foarte) generale sunt, de asemenea, nesatisfăcătoare întrucât sunt prea opace. Chiar dacă ni s-ar prezenta cel mai solid argument pentru o legătură între șomaj și războaiele de agresiune, precum și un argument persuasiv că s-au controlat toți „factorii terți" plauzibili, tot am vrea să știm *cum* cauzează șomajul războaie de agresiune. Putem chiar crede că explicația este corectă și să nu fim mulțumiți de ea. După cum am arătat într-un capitol anterior, acesta a fost statutul explicațiilor care se bazau pe legea gravitației înainte de teoria relativității generale. Acțiunea

la distanță era un lucru atât de misterios, încât mulți refuzau să creadă că lucrurile se vor opri aici. Dar, întrucât legea permitea efectuarea de predicții corecte până la câteva zecimale, scepticii trebuiau să accepte că lucrurile se întâmplă „ca și cum" legea ar fi corectă, chiar dacă nu puteau accepta existența unei forțe care poate „acționa acolo unde nu este prezentă".

Mecanismele

Sunt convins că cititorii vor fi ajuns deja la concluzia că exemplele de așa-zise legi universale din exercițiul de mai sus sunt destul de implauzibile. De acord. În parte, lipsa de plauzibilitate se poate datora lipsei mele de imaginație în elaborarea exemplelor, dar cred că există și motive mai profunde. Există foarte puține legi generale bine statuate în științele sociale. „Legea cererii" – când prețurile urcă, consumatorii cumpără mai puțin – este bine susținută de fapte, dar destul de nesatisfăcătoare dacă o gândim ca lege.[1] Legea gravitației, de exemplu, nu spune doar că, pe măsură ce distanța dintre două obiecte crește, forța de atracție dintre ele scade, ci ne spune cu cât scade (invers proporțional cu pătratul distanței). Nu există ceva de felul legii gravitației în științele sociale.[2]

Legea cererii și legea lui Engel, conform căreia ponderea din venituri care este cheltuită pe mâncare scade pe măsură ce venitul crește, sunt așa-zise *legi slabe*. Pentru orice schimbare (în sus sau în jos) a variabilei independente, ele ne permit să prezicem *direcția* sau semnul schimbării (în sus sau în jos) pentru variabila dependentă. Aceste legi nu ne permit, însă, să prezicem *magnitudinea* schimbării. Cu toate că sunt slabe, astfel de legi au o anumită consistență, din moment ce ne permit să excludem o gamă largă de valori posibile ale variabilei dependente. Ele nu ne ajuta, însă, să specificăm valoarea precisă care va fi instanțiată în cadrul valorilor care nu au fost excluse.

Legea cererii nu este doar slabă, ci și foarte nepotrivită pentru scopuri explicative. După cum am văzut în capitolul 1, ea este compatibilă cu un anumit număr de presupoziții despre comportamentul consumatorilor. Pentru *a explica* de ce consumatorii cumpără mai puțin dintr-un bun atunci când acesta se scumpește ar trebui

---

[1] Mai mult, pentru anumite bunuri cererea crește odată cu prețul. Consumatorii pot fi atrași de un bun tocmai pentru că este scump („efectul Veblen") sau pot cumpăra mai puțin dintr-un bun dacă prețul scade, de exemplu, pâine, întrucât își pot cumpăra bunuri de calitate superioară, ca, de exemplu, carne („efectul Giffen").

[2] Trebuie să aduc o clarificare aici: se susține adesea că puterea sentimentelor altruiste față de alții variază invers proporțional cu distanța socială față de agent. Cu toate acestea, ideea de „distanță socială" este mai mult o metaforă, și nu un concept, și, oricum, formularea „variază invers proporțional" este mult mai imprecisă decât „variază invers proporțional cu pătratul distanței".

să adoptăm și să testăm o presupoziție specifică despre reacția consumatorilor individuali la modificările de preț. Cuvântul-cheie aici este „individual". În științele sociale, o explicație satisfăcătoare trebuie să fie, în cele din urmă, ancorată în ipoteze despre comportamentul individual. Acest principiu – cunoscut drept „individualismul metodologic" – este premisa cărții de față ca întreg. Acest principiu presupune că psihologia și, poate, biologia trebuie să aibă o importanță fundamentală în explicarea fenomenelor sociale. Ezitarea mea cu privire la biologie nu vine din faptul că o consider incapabilă în principiu să explice aspecte ale comportamentului uman, ci din acela că este o știință încă prea puțin evoluată pentru a face acest lucru.

Pentru a explica comportamentul individual trebuie să ne bazăm în bună măsură pe ceea ce numesc eu *mecanisme*. În mare, mecanismele sunt *modele cauzale care apar frecvent și sunt ușor de recunoscut, fiind declanșate în condiții în general necunoscute sau care au consecințe nedeterminate*. Mecanismele ne dau posibilitatea să explicăm, nu să prezicem. S-a susținut, de exemplu, că pentru fiecare copil care devine alcoolic, ca reacție la un mediu dominat de alcoolici, alt copil evită complet alcoolul ca reacție la același mediu. Ambele reacții presupun mecanisme: faptul de a face ceea ce fac părinții și opusul a ceea ce fac aceștia. Nu putem să ne pronunțăm din timp ce va deveni copilul unui alcoolic, dar, în cazul în care acesta ajunge fie abstinent, fie alcoolic, putem estima că știm de ce.

Nu susțin că aici operează un soi de indeterminare obiectivă; acest concept are prea puțin sens în afara mecanicii cuantice. Susțin doar că, de multe ori, putem explica un comportament dacă arătăm că acesta este o instanțiere a unui model cauzal general, chiar dacă nu putem explica de ce se produce acel model. Mecanismele conformării (de exemplu, să faci ce fac părinții tăi) și ale nonconformismului (să faci opusul a ceea ce fac părinții) sunt, ambele, foarte generale. Dacă putem arăta că, de fapt, comportamentul unui copil cu un părinte alcoolic este o instanță a unui mecanism sau a altuia, am dat o explicație a comportamentului său. S-ar putea obiecta că, din moment ce nu am arătat de ce copilul devine (să zicem) alcoolic, mai degrabă decât abstinent, atunci nu am explicat nimic. Sunt, desigur, de acord că a da seamă de evoluția spre un rezultat, și nu spre celălalt, constituie o explicație mai bună și nu mă îndoiesc că, în anumite cazuri, chiar putem face acest lucru. Dar a subsuma o instanță individuală unui model cauzal general înseamnă, de asemenea, a oferi o explicație. A ști că un copil a devenit alcoolic ca rezultat al conformismului înseamnă a elimina o parte din opacitatea rezultatului, chiar dacă nu pe toată, dat fiind că nu am explicat și de ce a urmat copilul calea conformistă.

Am spus că un mecanism „apare des și este ușor de recunoscut". Înțelepciunea populară a identificat un număr mare de asemenea modele.[1] În definiția mea preferată, „Proverbul este transmis de-a lungul mai multor generații. Acesta rezumă, într-o

[1] După cum vom vedea în capitolul 10, proverbele nu sunt întotdeauna înțelepte.

formulare succintă, un principiu general sau o situație obișnuită, iar când îl rostești, toată lumea înțelege ce vrei să spui." Mai mult, proverbele enunță de multe ori mecanisme (în sensul utilizat aici), și nu legi generale. Putem să ne gândim, de exemplu, la o tendință frapantă a proverbelor de a se prezenta în perechi mutual exclusive. Pe de o parte, avem „Mai răruț, dar mai drăguț" și, pe de altă parte, avem „Ochii care nu se văd se uită". Pe de o parte, ne putem gândi că fructul oprit are gustul cel mai bun, iar, pe de alta, că strugurii la care nu putem ajunge sunt acri. Pe de o parte, „Cine se aseamănă se adună", iar, pe de altă parte, „Extremele se atrag". Pe de o parte, avem „Așchia nu sare departe de trunchi", iar, pe alta, „Părinte rău, fiu risipitor". Pe de o parte, „Graba strică treaba", dar, pe de altă parte, „Bate fierul cât e cald". Pe de o parte, „Reamintirea unui necaz înseamnă a-l mai trăi o dată", dar, pe de altă parte, „Reamintirea pericolelor trecute este plăcută". (După cum voi remarca ceva mai încolo, ultimele două proverbe nu sunt mutual exclusive.) Și exemplele pot continua.

Multe perechi de mecanisme opuse par a nu fi captate de proverbe. Să luăm ca exemplu perechea efect de răspândire – efect de compensare. Dacă o persoană care muncește mult la serviciu merge în vacanță, ne-am aștepta, oare, ca aceasta să păstreze același ritm frenetic și în activitățile de relaxare (efect de răspândire) sau, dimpotrivă, ne-am aștepta să se relaxeze în totalitate (efect de compensare)? Ne-am aștepta, oare, ca într-o democrație cetățenii să fie înclinați spre sentimente religioase sau să aibă o aversiune față de religie? Dacă obișnuința de a decide pentru sine pe care o exercită în sfera politică se răspândește în sfera religioasă (efect de răspândire), ne-am aștepta la o credință religioasă insuficientă. Dacă lipsa unei autorități superioare în sfera politică îl determină să caute autoritatea în altă parte (compensare), atunci un regim politic democratic tinde mai degrabă să favorizeze religia. O întrebare contemporană, rămasă încă fără răspuns, este dacă violența la televizor stimulează violența în viața reală (efect de răspândire) sau o atenuează (efect de compensare).

Mecanisme asemănătoare se pot aplica și relațiilor dintre indivizi. Să luăm în considerare chestiunea explicării donațiilor la cutia milei. Cineva poate fi preocupat de eficiența donației. Dacă ceilalți dau puțin, donația sa are șanse să aibă un impact mai mare, prin urmare, este mai probabil să dea bani; dacă ceilalți dau mult, atunci donația sa contează mai puțin și e posibil ca acesta să nu mai dea deloc. Altcineva poate fi preocupat mai mult de echitate (între cei care donează). Dacă alții dau puțin, atunci persoana respectivă nu poate înțelege de ce ar trebui să dea ea mai mult; invers, dacă alții dau mult, acea persoană se poate simți obligată să le urmeze exemplul. Aceeași pereche de mecanisme se poate aplica și în situațiile de acțiune colectivă. Pe măsură ce o mișcare populară crește, unii dintre membri pot renunța la afiliere întrucât nu li se pare că au o importanță prea mare, în timp ce alții ar putea să adere întrucât nu li se pare corect să stea pe margine, în timp ce alții suportă toate costurile (capitolul 24).

Chiar și proverbele care nu își găsesc o pereche opusă exprimă de cele mai multe ori mecanisme, și nu legi. Ar fi absurd dacă am interpreta proverbul „Cei mai buni înotători se îneacă cel mai des" în sensul că predispoziția spre înec crește odată cu abilitatea de a înota. Totuși, în cazul unor înotători încrederea în abilitățile lor se dezvoltă mai rapid decât abilitățile în sine, lucru care îi poate face să își asume riscuri cărora nu le pot face față („Înaintea prăbușirii merge trufia și semeția înaintea căderii."). Să luăm în considerare un proverb la care mă voi mai referi pe parcursul cărții, „Credem cu ușurință ceea ce sperăm și ceea ce ne provoacă frică"[1]. Chiar dacă proverbul nu este plauzibil ad litteram, ca lege universală ne aduce aminte că, în plus față de bine-cunoscutul fenomen al gândirii dezideraative, se mai află și o înclinație mai puțin înțeleasă, pe care o putem numi *gândire demotivantă*.[2] În cele din urmă, să luăm în considerare aserțiunile proverbiale „Copilul cu multe moașe rămâne cu buricul netăiat" și „Prea mulți bucătari strică ciorba". Din nou, valoarea proverbelor nu stă în aceea că enunță vreo lege universală, ci în faptul că sugerează mecanisme. Primul proverb poate fi adevărat dacă fiecare moașă crede că celelalte sunt la datorie (să ne amintim de cazul lui Kitty Genovese), iar al doilea este adevărat dacă fiecare bucătar crede că niciunul dintre colegii lui nu a pus sare în ciorbă.

Atunci când am definit noțiunea de mecanism am spus, de asemenea, că mecanismele sunt „declanșate în condiții necunoscute sau cu consecințe nedeterminate". Cele mai multe mecanisme care se regăsesc în proverbele pe care le-am citat se subsumează primei categorii. Nu știm ce condiții vor declanșa conformismul sau nonconformismul, gândirea dezideraativă sau demotivantă, preferința adaptivă (strugurii acri) sau preferința contraadaptivă (iarba e mai verde). Știm că cel mult unul dintre membrii fiecărei perechi se va realiza, dar nu ne putem da seama care. Este foarte importantă precizarea „cel mult", întrucât există posibilitatea ca unii oameni să nu se supună vreunuia dintre mecanismele-pereche. Autonomia autentică presupune a nu fi nici conformist, nici nonconformist. Credințele oamenilor pot fi independente de dorințele lor și dorințele lor pot fi independente de oportunități.

---

[1] Cu toate că acest proverb nu enunță o pereche de propoziții contrare, ne putem imagina enunțuri proverbiale cu un sens echivalent lui „credem cu ușurință ceea ce sperăm" și „suntem foarte neîncrezători în ceea ce sperăm". De fapt, suntem de cele mai multe ori ne situăm, într-un mod părtinitor, împotriva a ceea ce sperăm deoarece ne ferim excesiv de gândirea dezideraativă.

[2] Ideea poate fi precizată într-un mod mai general, astfel încât să includă atât faptul de a nu crede ceea ce speri, cât și faptul de a crede ceea ce provoacă frică. În nota de subsol precedentă, am prezentat un mod de a obține prima variantă. Cât o privește pe a doua, poate că dorința de a te pregăti pentru ce e mai rău tinde să întărească convingerea că răul se va produce. Simplul proces de a parcurge un scenariu mental îl poate întipări pe acesta în minte și îl poate face să pară nu doar vag posibil, ci chiar plauzibil sau probabil.

În alte cazuri, proverbele sugerează declanșarea simultană a două mecanisme cu efecte opuse în raport cu rezultatul. În acest caz, indeterminarea se regăsește în evaluarea *efectului net* al mecanismelor, mai degrabă decât în stabilirea mecanismului care se va declanșa (dacă e cazul). Să luăm în considerare următorul exemplu: „Nevoia te învață" și „E scump să fii sărac". Primul proverb afirmă o legătură cauzală între sărăcie și o *dorință* arzătoare pentru inovație, al doilea afirmă o legătură între sărăcie și foarte puține *oportunități* pentru inovație. Întrucât comportamentul este modelat atât de dorințe, cât și de oportunități (capitolul 9), în general nu putem să ne dăm seama dacă impactul net al sărăciei asupra inovației este favorabil sau nefavorabil. Ori să luăm în considerare perechea de proverbe menționată mai devreme, „Reamintirea unui necaz înseamnă a-l mai trăi o dată" *versus* „Reamintirea pericolelor trecute este plăcută". Primul proverb se bazează pe așa-zisul „efect al împroprietăririi": amintirea unei experiențe neplăcute este în sine o experiență neplăcută.[1] Al doilea proverb se bazează pe un „efect al contrastului": amintirea unei experiențe neplăcute sporește valoarea prezentului.[2] În general, nu ne putem da seama dacă efectul net al unei experiențe anterioare neplăcute va fi favorabil sau nefavorabil bunăstării ulterioare.

Din nou, nu trebuie să ne rezumăm la proverbe. Putem lua în considerare două mecanisme neproverbiale implicate în așa-numita „psihologie a tiraniei". În cazul în care un tiran sporește oprimarea supușilor săi, sunt probabile două efecte. Pe de o parte, pedepsele sporite vor preveni rezistența sau răscoala supușilor. Pe de altă parte, pe măsură ce conducătorul se comportă din ce în ce mai mult ca un tiran, ura supușilor săi va crește. Ca orice asupritor, și acesta va inspira în același timp atât ură, cât și frică. Dacă ura domină frica, asuprirea va avea un efect opus celui scontat. În țările aflate sub ocupație nazistă în al Doilea Război Mondial, membrii Rezistenței exploatau uneori acest mecanism atunci când ucideau soldați germani pentru a provoca represalii, bazându-se pe presupunerea că efectul descurajator va fi dominat de „efectul tiranic"[3]. Se poate aduce în discuție și cazul, oarecum asemănător, al unei persoane care întâm-

---

[1] Invers, reamintirea unei experiențe plăcute este în sine o experiență plăcută. După cum spune Tennyson: „Mai bine să iubești și să pierzi decât să nu fi iubit deloc."

[2] Invers, reamintirea unei experiențe plăcute din trecut face prezentul să își piardă din valoare. După cum spune Donne: „E mai bine să fii slut decât să fi fost vreodată frumos."

[3] Uneori ura și-a găsit o altă țintă. În trei sate din Italia Centrală și de Nord, naziștii au fost autorii unor represalii sălbatice în 1944, ceea ce i-a făcut pe săteni să le fie ostili partizanilor din Rezistență chiar și cincizeci de ani mai târziu întrucât aceștia erau autorii indirecți sau poate chiar autorii „adevărați" ai masacrului. Atunci când A cauzează uciderea de către B a lui C, rudele și prietenii lui C își pot îndrepta mânia mai degrabă spre A decât spre B. În rândul mișcărilor de Rezistență care se opuneau trupelor de ocupație nazistă se pot observa ambele tipuri de mecanisme.

pînă o anumită barieră sau impediment în atingerea unui scop. Această amenințare la adresa libertății de acțiune poate determina ceea ce psihologii numesc „reactanță" – o anumită motivare în vederea recuperării sau restabilirii libertății. Efectele barierei asupra reactanței sunt opuse și, în general, nu ne putem da seama care efect va predomina.[1] Acest aspect poate fi ilustrat gândindu-ne la efectul provocat de părinții care ascund de băiețelul lor o tobă pe care acesta o bate cu plăcere, dar care pe părinți îi deranjează.[2]

Chiar și atunci când cunoaștem efectul net, este posibil să nu fim în măsură să îl explicăm. Să presupunem că avem posibilitatea să observăm și să măsurăm un efect net zero al efectelor de împroprietărire și de contrast cu privire la o experiență plăcută în trecut. Acest rezultat se poate produce în două moduri. Cu toate că meniul franțuzesc de trei stele de care m-am bucurat anul trecut a redus plăcerea pe care mi-au produs-o experiențe ulterioare în restaurante franțuzești obișnuite, acest efect defavorabil asupra bunăstării mele este contrabalansat în totalitate de amintirea meniului excepțional de care am avut parte. Și totuși, observarea unui efect net zero este, de asemenea, perfect consistentă atât cu efectul zero de împroprietărire, cât și de contrast, dar și cu ambele efecte, la fel de puternice. Cât timp nu știm cum stau lucrurile, nu putem pretinde că am explicat rezultatul. Pentru a evalua tăria fiecărui efect am putea să ne uităm la rezultat într-o situație în care celălalt rezultat potențial nu este așteptat să se producă. Dacă, după cum pare plauzibil, plăcerea pe care mi-o oferă bucătăria grecească nu este afectată de meniul franțuzesc de trei stele, atunci putem identifica strict tăria efectului de împroprietărire.

O nedeterminare înrudită poate apărea cu privire la primul tip de mecanisme, cele care sunt declanșate în „condiții în general necunoscute". Să luăm în considerare, din nou, cazul părinților alcoolici. Dacă ne raportăm la întreaga populație de alcoolici (sau la un eșantion reprezentativ), să presupunem că, în medie, copiii acestora nu beau nici mai mult, nici mai puțin în comparație cu copiii acelora care nu sunt alcoolici. Pentru simplificarea lucrurilor, vom lăsa deoparte influența factorilor genetici. Vom putea înțelege această descoperire ipotetică în două moduri. Pe de o parte, copiii alcoolicilor nu trebuie neapărat să fie nici conformiști, nici nonconformiști: cu alte cuvinte, aceleași cauze care formează comportamentul copiilor cu părinți nealcoolici contribuie la formarea comportamentului copiilor cu părinți alcoolici. Pe de altă

---

[1] O trăsătură specială a acestui exemplu este aceea că unul dintre cele două efecte competitoare (reactanța) este indus de celălalt efect (bariera). În celelalte exemple, cele două efecte sunt produse simultan de o cauză comună (de exemplu, asuprirea tiranului).

[2] Putem ilustra reactanța și printr-un alt exemplu: copiii (dar și adulții) pot refuza orice fel de propunere despre modul în care să-și petreacă ziua, *chiar și alternativa pe care o preferă*, întrucât nu vor să simtă că li se impune ceva.

parte, există posibilitatea ca o jumătate dintre copiii alcoolicilor să fie conformiști, iar cealaltă jumătate să fie nonconformiști, ceea ce duce la un efect net egal cu zero.

Similar, teoriile referitoare la comportamentul de vot au identificat atât un mecanism de susținere a candidatului inferior, cât și un mecanism de susținere a aceluia mai tare. Votanții din prima categorie au tendința să voteze candidatul situat pe locuri inferioare în sondaje, pe când cei din categoria a doua votează pentru candidatul cu cele mai mari șanse de câștig. Dacă avem o distribuție egală a celor două tabere, e posibil să nu observăm niciun efect net, prin urmare, sondajele ar fi un bun instrument de predicție al votului real. Totuși, chiar dacă nu sesizăm o influență a sondajelor asupra votului raportat la nivelul masei agregate de votanți, acest lucru nu indică faptul că indivizii în sine nu sunt influențați de sondajele de opinie. Efecte agregate slabe ale violenței prezentate la televizor asupra vieții reale pot masca efecte foarte puternice, dar opuse, la nivelul subgrupurilor. În toate aceste cazuri, un agregat neutru poate reflecta fie o populație omogenă de indivizi neafectați, fie o populație eterogenă de indivizi care sunt afectați fiecare în parte, dar în direcții opuse. Nevoia de a elimina această ambiguitate ne oferă un nou argument pentru individualismul metodologic. Trebuie să ne raportăm la comportamentul indivizilor în sine pentru a putea să explicăm comportamentul la nivel agregat.

Mecanisme moleculare

Am luat în considerare ceea ce am putea numi mecanisme „atomice": reacții psihologice elementare care nu pot fi reduse la alte mecanisme de același nivel. De asemenea, ne putem pune întrebarea cât de departe ne vor duce aceste mecanisme psihologice în explicarea fenomenelor sociale. Răspunsul stă în aceea că putem folosi mecanismele atomice ca pe niște cărămizi în construcția unor mecanisme „moleculare" mai complexe. Din nou, putem începe demersul făcând apel la proverbe. De exemplu: „Frica este adesea mai mare decât pericolul în sine" și „Frica sporește pericolul"; luate în considerare împreună, duc cu gândul la faptul că frica excesivă poate să își creeze propria justificare. Un proverb englezesc spune că „În fiecare turmă există o oaie neagră". Un proverb franțuzesc ne spune că „E nevoie de o singură oaie rea pentru a distruge o turmă". Dacă le considerăm împreună, putem infera că fiecare turmă se va distruge.[1]

Să luăm în considerare un corp guvernat autonom, cum ar fi un departament dintr-o universitate sau o cooperativă muncitorească. Următorul scenariu este întâlnit suficient de des și este destul de inteligibil pentru a fi considerat mecanism molecular.

[1] Am formulat într-un mod mai liber aceste proverbe. În înțelesul său literal, sensul expresiei franceze *une brebis galeuse* se referă la o oaie cu o boală de piele cauzată de un parazit arahnid.

În primul rând, prin legile probabilității, într-un grup de douăzeci de indivizi sau mai mare se va afla probabil cel puțin un „Gică-contra", care este împotriva tuturor, unul care „încurcă lucrurile" și care poate fi definit drept „cineva care afectează șansele de victorie ale unui adversar, fără să fie un potențial învingător". În al doilea rând, într-un grup care conține un astfel de personaj guvernarea colectivă este foarte dificilă. Discuțiile durează la nesfârșit, decizii deja luate sunt permanent puse în discuție, colegialitatea cu caracter informal este înlocuită de un aer formal, apare ranchiuna și așa mai departe. Într-un final, grupul va primi cu bunăvoință guvernarea unui comitet executiv mai restrâns sau poate chiar va prefera ca puterea să se afle în mâinile unui singur individ.

Lăsând proverbele la o parte, să luăm în considerare un alt mecanism molecular. De secole sau chiar milenii, elitele au fost precaute în ceea ce privește democrația ca formă de guvernare întrucât se credea că aceasta ar permite tot felul de comportamente periculoase și licențioase. Și totuși, simpla oportunitate a unui comportament periculos nu va declanșa în sine un astfel de comportament: trebuie să existe și un motiv. Oare democrațiile înăbușă cumva pornirile firești ale cetățenilor, astfel încât aceștia să facă doar ceea ce le permite regimul politic? Cel puțin, aceasta susținea Tocqueville. El credea că, pentru a satisface nevoia de autoritate pe care politica nu o asigură, cetățenii dintr-o democrație vor apela la religie, care are tendința de a limita și înfrâna dorințele cetățenilor. Criticii democrației nu au înțeles cum stau lucrurile, susținea el, întrucât s-au concentrat doar pe oportunități, neglijând, în același timp, dorințele. Cu toate că Tocqueville a enunțat argumentul acesta ca și cum ar reda o lege universală, este mult mai plauzibilă înțelegerea sa în termenii mecanismelor. În primul rând, dacă efectul care funcționează este acela al răspândirii, și nu acela al compensării, atunci lipsa unei autorități politice va slăbi religia, nu o va întări. În al doilea rând, chiar dacă funcționează efectul răspândirii, nu putem trage nicio concluzie privitoare la efectul net. Dacă setul de oportunități este suficient de mare și dorințele sunt doar slab înăbușite, efectul net al democrației mai degrabă va spori incidența respectivului tip de comportament.

Cele două perechi de mecanisme sunt reprezentate succint în Figura 2.1. *Dacă* influența democrației asupra religiei este mediată de efectul de compensare, și nu de

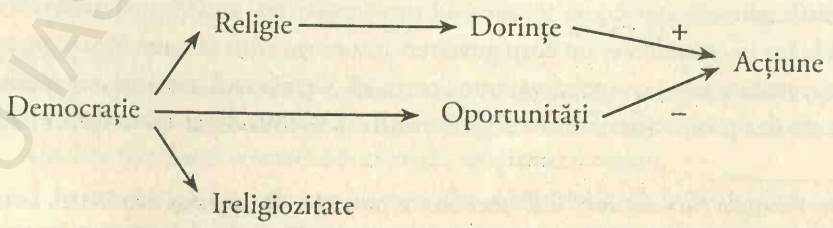

FIGURA 2.1

cel de răspândire, atunci societățile democratice vor fi religioase. *Dacă* efectul negativ al democrației asupra dorințelor (mediat de religie) este suficient de puternic pentru a contrabalansa efectul pozitiv al democrației asupra oportunităților, atunci cetățenii din societățile democratice vor avea un comportament moderat.[1]

Mecanisme și legi

Deseori, cea mai bună explicație pe care o putem da este aceea care face apel la mecanisme. Uneori, însă, suntem în măsură să facem ceva în plus. Odată ce am identificat faptul că un mecanism este „declanșat în condiții în general necunoscute", putem identifica și condițiile declanșatoare. În acest caz, mecanismul va fi înlocuit de o lege, chiar dacă o lege de obicei slabă, în sensul definit mai devreme.

Bunul-simț ne-ar spune că un dar îl face fericit pe acela care îl primește. În caz contrar, vina aparține aceluia care îl primește. Moraliștii clasici – de la Montaigne până la La Bruyère – au susținut că primirea unui dar provoacă mai degrabă resentiment, decât recunoștință. Se pare că atât simțul comun, cât și moraliștii clasici sunt pe urmele a ceva, dar nu pot să ne spună când ne putem aștepta la un rezultat și când la celălalt. Un moralist din Antichitatea clasică, Publilius Syrus, a enunțat și factorii declanșatori: un dar modest creează o obligație, un dar scump creează un dușman.[2] Făcând apel la dimensiunea de factor declanșator a darului, am transformat perechea de mecanisme într-un enunț (oarecum) legic.[3] Putem menționa un alt exemplu: am putea fi în măsură să enunțăm când anume poate fi rezolvată o tensiune între o dorință și o credință („disonanță cognitivă") prin modificarea credinței și când poate fi rezolvată prin modificarea dorinței.[4] Credințele cu conținut strict factual nu pot fi schimbate prea ușor (capitolul 7). Persoana care a plătit 75 de dolari pentru un bilet la un spectacol pe Broadway nu se poate auto-amăgi prea ușor că ar plăti doar 40. Cu

---

[1] În plus, Tocqueville a susținut că tirania majorității va exercita o influență moderatoare. „Unele guverne caută să salvgardeze bunele moravuri prin condamnarea autorilor de cărți licențioase. În Statele Unite, nimeni nu este condamnat pentru o astfel de lucrare, dar nici nu este nimeni tentat să scrie așa ceva."

[2] Trișez puțin aici ca să-mi iasă exemplul, din moment ce Syrus se referă la împrumuturi, nu la daruri. Deși pare plauzibil că atât împrumutul, cât și darul care constă într-o sumă mare de bani poate să îl facă pe cel care primește să se simtă ofensat, probabil că acest lucru se întâmplă în moduri diferite în cele două cazuri.

[3] Doar „întrucâtva", din moment ce mărimea precisă la care recunoștința se transformă în indignare va fi puternic dependentă de context. Aceeași remarcă se aplică și celorlalte cazuri discutate în text.

[4] Să ne aducem aminte, totuși, că tensiunea poate rămâne nerezolvată.

toate acestea, va putea să găsească anumite aspecte plăcute ale spectacolului și se va convinge că acestea sunt mult mai semnificative decât cele mai puțin plăcute.

Mai devreme m-am referit la contrastul dintre mecanismul „fructului oprit" și acela al „strugurilor acri". În unele cazuri putem fi în măsură să prezicem care dintre cele două mecanisme se va declanșa. Într-un experiment, unor subiecți aflați într-o anumită stare li s-a cerut să facă un top al preferințelor pentru patru cântece și li s-a spus că a doua zi vor primi o înregistrare cu unul dintre ele, ales la întâmplare. Altor subiecți, în altă context, li s-a spus, după ce au făcut respectivul top, că vor fi lăsați să își aleagă înregistrarea cu melodia preferată. A doua zi, tuturor subiecților li s-a spus că discul numărul trei în ordinea preferințelor este indisponibil și li s-a cerut să facă un nou top, ca parte a tentativei de a descoperi felul în care o a doua audiție influențează evaluarea. Conform predicțiilor făcute prin teoria reactanței, subiecții aflați în prima stare au manifestat efectul „strugurilor acri", declasând valoarea opțiunii indisponibile, în timp ce al doilea grup a manifestat efectul „fructului oprit", mărind valoarea opțiunii indisponibile. (Grupul de control, căruia nu i s-a comunicat nimic despre eliminarea unei opțiuni, nu a manifestat nicio schimbare.) Diferența crucială stă în aceea că primul grup a avut parte de experiența unei amenințări la adresa libertății lor, pe când al doilea grup nu.

Să tratăm acum un exemplu ceva mai complex. În legătură cu perechea de proverbe „Mai răruț, dar mai drăguț" și „Ochii care nu se văd se uită", există și al treilea proverb, care sugerează un factor declanșator: „O scurtă absență poate face mult bine." La Rochefoucauld a propus un factor diferit: „Absența micșorează pasiunile moderate, dar le stârnește pe cele mari, așa cum vântul stinge o lumânare, dar aţâță un foc." Aceste propuneri plauzibile nu constituie legi foarte puternice. Pentru a fi în măsură să facem predicții ale evoluției unei anumite pasiuni ar trebui să cunoaștem dimensiunea unei absențe scurte (trei săptămâni?), precum și a unei pasiuni puternice (una care te ține treaz noaptea?). De asemenea, ar trebui să specificăm modul în care interacționează durata absenței și puterea pasiunii, astfel încât să genereze creșterea sau descreșterea unei pasiuni în timpul unei absențe. Să o cercetăm pe ultima dintre aceste probleme.

Interacțiunea dintre cauze

În general, științele sociale nu reușesc să explice prea bine modul în care cauzele interacționează pentru a produce efecte comune. Cel mai adesea, se presupune că fiecare cauză contribuie separat la efect (un „model aditiv"). Pentru a explica nivelul venitului, de exemplu, am putea presupune că acesta este cauzat parțial de venitul părinților și parțial de educația părinților, iar apoi se utilizează metode statistice pentru a determina contribuția relativă a acestor două cauze. Pentru exemplul pe care l-am pus

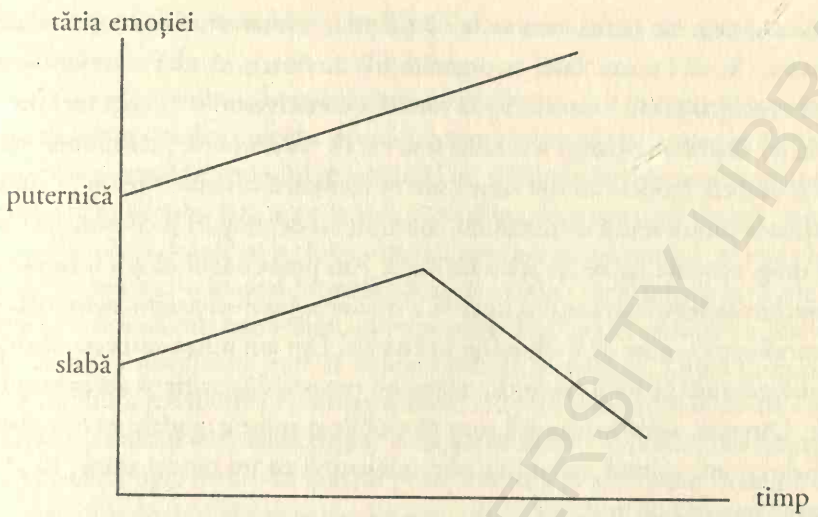

FIGURA 2.2

în discuție, acest mod de abordare s-ar putea să nu fie potrivit. Durata absenței ar putea să nu constituie o contribuție separată la intensitatea emoției resimțite în perioada de după absență; efectul ei ar putea depinde mai degrabă de puterea emoției de dinaintea absenței. Acest efect al interacțiunii este prezentat în Figura 2.2.

Totuși, există oameni de știință care susțin că lumea – sau cel puțin acea parte a ei care constituie domeniul lor de studiu – pur și simplu nu manifestă prea multe interacțiuni de acest tip. Se întâmplă foarte rar, susțin ei, ca, pentru niveluri joase ale variabilei independente X, variabila Z să crească (descrească) împreună cu variabila dependentă Y, în timp ce, pentru niveluri mari ale lui X, o creștere la nivelul lui Y va cauza o creștere (descreștere) a lui Z. Relația ipotetică din Figura 2.2 ar fi o excepție (în cazul în care există). Cel mult, susțin aceștia, ceea ce se întâmplă este că, la niveluri joase ale lui X, Y are un efect restrâns asupra lui Z, în timp ce efectul se face simțit la niveluri mai mari ale lui X. În explicarea venitului, de exemplu, se poate presupune că venitul părinților contribuie mai mult sau mai puțin la diferite niveluri ale educației părinților. Acest tip de interacțiune poate fi exprimat printr-un termen multiplicativ al interacțiunii, astfel încât Z devine o funcție de X, Y și XY. Prin contrast, *inversarea* efectului cauzal al lui Y asupra lui Z la niveluri mari ale lui X nu poate fi exprimat în acest mod. Dacă putem crede afirmația că astfel de inversări sunt rare, atunci nu trebuie să ne facem prea multe griji în privința lor.

Existența unui efect de interacțiune se poate supune aceluiași tip de indeterminare pe care îl regăsim la nivelul mecanismelor în general. Să ne oprim asupra interacțiunii dintre vârstă și atitudinile politice de bază, considerate drept cauze ale extremismului. Am putea bănui că organizațiile de tineret vor avea o orientare mai de stânga decât

partidele din care fac parte, ceea ce le dă tinerilor conservatori o tentă mai deschisă de albastru.[1] Se mai poate, însă, ca organizațiile de tineret să aibă o atitudine mai extremistă decât partidele – ceea ce le dă tinerilor conservatori o nuanță mai închisă de albastru, iar tinerilor socialiști o nuanță mai vie de roz. Ambele presupuneri sunt plauzibile și ambele modele au fost observate în realitatea curentă. Putem să studiem și interacțiunea dintre starea de dinaintea consumului de droguri și consumarea efectivă a unui drog, socotite cauze ale stării de după. Am putea bănui că administrarea unui drog precum alcoolul sau cocaina duce la o *creștere a bunei-dispoziții*, atenuează depresia și transformă o stare de mulțumire în euforie. Dar am putea suspecta că drogurile sunt *multiplicatori* ai stării de spirit, adâncind proasta-dispoziție și accentuând starea de bine. Din nou, ambele bănuieli sunt plauzibile și ambele modele au fost observate. În ambele cazuri, primul mecanism este compatibil cu un model aditiv, iar al doilea presupune un efect de inversare.

Atunci când suntem puși în fața unor date contradictorii, avem la dispoziție mai multe răspunsuri, în afară de adăugarea unui termen de interacțiune sau a unuia care să ajusteze o curbă empirică. Când facem apel la o ajustare a curbei empirice, menținem variabilele independente și dependente la aceeași valoare și încercăm să găsim o funcție matematică prin care să dăm seama de deviațiile statistice. Într-un exercițiu de extragere a cunoștințelor din date, funcția matematică este menținută fixă (de obicei, un model aditiv simplu) și se caută variabilele independente care se potrivesc cu variabilele dependente. Să presupunem că printr-o „bună potrivire" înțelegem o corelație care ar avea o probabilitate de doar 5% să apară întâmplător. În orice studiu al unui fenomen social complex, cum este venitul, se poate înregistra cu ușurință un număr semnificativ de variabile care l-ar putea influența.[2] De asemenea, există, probabil, destule moduri în care am putea conceptualiza ideea de venit. Ar fi foarte puțin probabil ca niciuna dintre variabilele independente să nu se coreleze cel puțin la un nivel de 5% cu una dintre definițiile pe care le dăm venitului.[3] Legile probabilității susțin

---

[1] Culoarea Partidului Conservator din Marea Britanie. (N. tr.)

[2] Astfel, într-un studiu longitudinal despre relația dintre practicile materne și rezultatul creșterii copilului, doar 35 din 552 de corelații au fost semnificative din punct de vedere statistic „la nivelul $p < 0,05$" (ceea ce înseamnă una din douăzeci de șanse ca rezultatul să apară din întâmplare).

[3] Teoria ar putea sugera că vremea rea îi deprimă pe cei care fac tranzacții la bursă, cauzând o încetare a tranzacțiilor. Oamenii de știință, însă, indică rezultatul opus când vremea rea este definită ca vreme cu cer complet acoperit de nori. Dacă se schimbă definiția astfel încât să avem doar un procent de 80% din cer acoperit de nori, semnul corelației se inversează în mod miraculos.

că cea mai improbabilă coincidență ar fi aceea ca niciodată să nu apară coincidențe improbabile.[1]

Odată ce a identificat o funcție matematică adecvată sau un set adecvat de variabile independente sau dependente, omul de știință poate căuta o poveste cauzală care să furnizeze o intuiție în sprijinul descoperirilor. Ordinea în care rezultatele urmează a fi prezentate publicului larg în reviste de specialitate este inversată de cele mai multe ori. Se susține că s-a pornit cu o teorie cauzală, apoi s-a căutat modul cel mai plauzibil de a o transforma într-o ipoteză formală, iar apoi s-a găsit confirmarea la nivelul datelor empirice.[2] Aceasta nu este știință, ci impostură. În științele naturii nu este nevoie de nicio „logică a justificării" care să corespundă sau să reflecte „logica descoperirii". Odată ce s-a enunțat o ipoteză în forma sa finală, este irelevant modul în care s-a ajuns la ea. Ceea ce contează sunt consecințele ei de acum încolo, nu rădăcinile acesteia. Iar acest lucru se întâmplă pentru că ipoteza poate fi testată pe un număr indefinit de observații, mai mult decât acelea care l-au inspirat inițial pe omul de știință care a propus ipoteza. În științele sociale (și în științele umaniste), cele mai multe explicații folosesc un set finit de date. Întrucât procedurile de colectare a datelor sunt de cele mai multe ori nestandardizate, se poate întâmpla ca oamenii de știință să nu fie în măsură să testeze ipotezele pe date noi.[3] Iar dacă procedurile *sunt* standardizate, datele pot eșua în

---

[1] Am două astfel de experiențe personale. Prima dată când am vizitat New Yorkul, mi-am cumpărat bilete pentru două spectacole pe Broadway, unul conceput în jurul muzicii lui Fats Waller, iar celălalt centrat pe muzica lui Duke Ellington. Mai erau puține bilete, așa că a trebuit să mă mulțumesc cu ce s-a găsit – în ambele cazuri am stat în rândul H, locul 130. Acest aspect a fost un pic straniu, dar altă coincidență mi s-a părut mai semnificativă. Există două experiențe pe care le-am trăit o singură dată. Prima este aceea de a fi fost invitat la un dineu, iar eu am uitat de invitație. A doua este aceea de a fi fost invitat la un dineu doar pentru a fi sunat de amfitrion cu jumătate de oră înainte pentru a mi se comunica anularea petrecerii pe motiv de boală. Coincidență care, pentru o clipă, m-a făcut să cred că sunt spionat de cineva pentru că era vorba despre una și aceeași petrecere.

[2] Prin urmare, există trei probleme legate de utilizarea corelației drept ghid al cauzalității. În primul rând, corelația poate apărea din pură întâmplare, fără să aibă o interpretare cauzală. În al doilea rând, corelația poate avea o interpretare cauzală indirectă dacă cele două fenomene corelate sunt efecte comune ale unui „factor terț". În al treilea rând, direcția cauzalității poate fi ambiguă.

[3] Am putea să evităm – sau măcar să mediem – această problemă dacă ne restrângem scopurile. Dacă avem la dispoziție un set suficient de mare de date, ne putem concentra mai întâi pe un eșantion reprezentativ și putem ignora restul. Odată ce am explicat cât putem de bine subsetul de observații, putem aplica explicația pe setul întreg de date, să vedem dacă ține. Dacă ține, este mult mai puțin probabil să fie o aparență. Un alt mod de a menține onestitatea oamenilor de știință ar fi ca revistele de specialitate să refuze să ia în considerare articole propuse spre publicare dacă ipotezele de testat, împreună cu procedurile pentru testarea lor, nu au fost

reflectarea unei realități în schimbare. Este imposibil să explici comportamentul consumatorilor fără să ții cont de apariția unor noi produse pe piață și de modificările de preț aduse produselor mai vechi.

Nu încape îndoială că practici de felul acesta există. Nu știu cât de des sunt întâlnite, ci doar că sunt suficient de răspândite încât să provoace îngrijorare în rândul experților din științele sociale. Cauza principală a problemei este, poate, faptul că înțelegem într-un mod neadecvat cauzalitatea multifactorială. Dacă am avea intuiții puternice despre modul în care mai multe cauze pot interacționa pentru a produce un anumit efect, atunci nu ar trebui să ne bazăm pe procedura mecanică de a „adăuga un termen de interacțiune" atunci când un model aditiv eșuează. Dar, întrucât intuițiile noastre sunt slabe, nu știm exact ce anume căutăm, iar improvizarea unor modele pare a fi singura alternativă – cel puțin dacă păstrăm țelul ambițios de a da explicații în termeni de legi. Însă, date fiind pericolele potențiale ale oricărei improvizații, poate ar fi cazul să ne propunem țeluri mai puțin ambițioase.

◈◈◈

Notă bibliografică

Multe dintre ideile discutate în acest capitol sunt adaptate din capitolul 1 al lucrării mele *Alchemies of the Mind* (Cambridge University Press, 1999). Acolo am citat și lucrări ale lui Raymond Boudon, Nancy Cartwright și Paul Veyne, care pledează pentru abordări asemănătoare. O abordare mai recentă se găsește în P. Hedström, *Dissecting the Social* (Cambridge University Press, 2005). Moduri utile de a aborda mecanismele psihologice se găsesc în F. Heider, *The Psychology of Interpersonal Relations* (Hillsdale, NJ, Lawrence Erlbaum, 1958), și R. Abelson, *Statistics as Principled Argument* (Hillsdale, NJ, Lawrence Erlbaum, 1995). Ultima lucrare pune la dispoziție considerații înțelepte și spirituale privitoare la capcanele analizelor statistice. O scurtă expunere-standard a ideii că știința explică prin apel la legi generale poate fi găsită în C. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, NJ, Prentice-Hall, 1966). Principiul individualismului metodologic este expus pe larg în Partea a IV-a din M. Brodbeck (ed.), *Readings in the Philosophy of the Social Sciences* (Londra, Macmillan, 1969) și în Partea a IV-a din M. Martin și L. McIntyre (editori), *Readings in the Philosophy of Social Science* (Cambridge, MA: MIT Press, 1994); vezi, de asemenea, și K. Arrow, „Methodological Individualism and social knowledge", *American Economic Review: Papers and Proceedings*, 84 (1994), pp. 1-9. Am abordat problema

---

predate editorului (să zicem) cu doi ani în avans. Ketil Bruun, un cercetător specializat în cercetarea alcoolului, a adoptat această practică.

proverbelor într-un mod mai sistematic în „Science et sagesse: Le rôle des proverbes dans la connaissance de l'homme et de la société", în J. Baechler (ed.), *L'acteur et ses raisons: Mélanges Raymond Boudon* (Paris, Presses Universitaires de France, 2000). Ideea „psihologiei tiranului" a fost preluată din J. Roemer, „Rationalizing revolutionary ideology", *Econometrica*, 53 (1985), pp. 85-108. Studiul subiecților cărora le-au fost promise discuri este J. Brehm *et al.*, „The attractiveness of an eliminated choice alternative", *Journal of Experimental Social Psychology*, 2 (1966), pp. 301-313. O introducere generală la teoria reactanței este prezentată de R. Wicklund, *Freedom and Reactance* (New York, Wiley, 1974). Scepticismul cu privire la interacțiunea care determină efecte inverse poate fi găsit în R. Hastie și R. Dawes, *Rational Choice in an Uncertain World* (Thousand Oaks, CA, Sage, 2001), capitolul 3. Povestea din nota de subsol despre 6% corelație semnificativă a apărut în R. R. McCrae și P. T. Costa, „The paradox of parental influence", în C. Perris, W. A. Arrindell și M. Eisemann (ed.), *Parenting and Psychopathology* (New York, Wiley), pp. 113-114. Exemplul redat într-una dintre notele de subsol despre impactul vremii urâte asupra acelora care fac tranzacții la bursă este luat din P. Kennedy, „Oh no! I got the wrong sign! What should I do?", *Journal of Economic Education*, 36 (2005), pp. 77-92, articol care conține comentarii utile privitoare la costurile (și beneficiile!) extragerii cunoștințelor din date în general.

BCU IASI CENTRAL UNIVERSITY LIBRARY

Capitolul 3
INTERPRETAREA

Interpretare și explicație

În multe lucrări dedicate științelor umaniste, *interpretarea*, și nu explicația, are un rol central. În tradiția germană, cel mai adesea se trage o linie de demarcație între „științele spiritului" (*Geisteswissenschaften*) și științele naturii (*Naturwissenschaften*). Pentru primele, ni se spune, procedura adecvată este cea a interpretării sau a „înțelegerii" (*Verstehen*). Limbajul adecvat pentru cele din urmă este acela al explicației (*Erklären*). Max Weber a scris, de exemplu, că științele naturii nu urmăresc să „înțeleagă" comportamentul celulelor.

Putem, astfel, să ne întrebăm dacă științele sociale se bazează pe înțelegere sau pe explicație. Eu cred că întrebarea este pusă greșit. După mine, a interpreta *înseamnă* a explica. Interpretarea nu este nimic altceva decât un caz special al metodei ipotetico-deductive (capitolul 1). Experții din domeniul științelor umaniste nu pot, de pildă, să folosească „empatia" ca pe o scurtătură în interpretarea comportamentului, din moment ce sentimentele empatice ale unui om de știință diferă de ale altuia. Pentru a decide între interpretări contrare aceste intuiții interpretative sau ipoteze (pentru că aceasta sunt, de fapt), este necesară o confruntare cu *experiența*. După cum am susținut în capitolul 1, experiența include nu doar faptele pe care încercăm să le înțelegem, ci și *fapte noi*, pe care altminteri nu ne-am fi gândit să le investigăm.[1]

Interpretarea este direcționată înspre acțiunea umană și spre produsul acțiunii umane, cum ar fi opera de artă. În capitolul 14 voi discuta interpretarea operelor literare, în special opere în care trebuie să înțelegem atât acțiunile personajelor, cât și alegerile făcute de autor. Această dublă chestiune nu apare în încercarea de a înțelege

---

[1] În științele experimentale, „faptele noi" pot fi fapte care sunt literalmente noi, cum ar fi cele generate de expunerea șobolanilor sau chiar a ființelor umane la condiții care nu există în natură. În științele umaniste și în cele non-experimentale, „nou" trebuie înțeles într-un sens epistemic, al „unui ceva neașteptat înainte", și nu în sensul ontologic, al „unui ceva inexistent înainte."

BUCHAREST UNIVERSITY LIBRARY

alte opere literare, precum și „arte fără cuvânt", cum ar fi pictura, sculptura sau muzica instrumentală. Totuși, chiar și în aceste opere, alegerile făcute de artist pot fi supuse unei analize similare cu cea pe care o voi propune pentru deciziile auctoriale. Artiștii fac alegeri în conformitate cu un criteriu al „îmbunătățirii" pe care poate că nici ei, nici noi nu suntem în măsură să îl formulăm în mod explicit, dar care este revelat în practică, atunci când artistul urmează o anumită schiță a unui proiect (sau a unui manuscris ori înregistrări) în favoarea alteia. Și totuși, relația dintre criteriul îmbunătățirii și psihologia umană este mult mai complicată și mai puțin înțeleasă în artele care nu implică limbajul, decât în literatura clasică. Însă nu voi încerca să mă ocup de asemenea probleme.

Raționalitate și inteligibilitate

În restul capitolului mă voi ocupa de interpretarea *acțiunii*. Atunci când vrem să interpretăm o acțiune, ni se cere să o explicăm în termenii credințelor și ai dorințelor (motivațiilor) anterioare ale agentului. În plus, ar trebui să explicăm aceste stări mentale în sine într-un mod care dă seamă de ele situându-le în cadrul complexului dorință-credință. O credință sau o dorință izolată care nu manifestă aceeași solidaritate normală cu alte stări mentale este un fapt brut care ne poate permite să explicăm un comportament, dar nu și să îl înțelegem.

Un mod paradigmatic de a explica acțiunea este de a demonstra că aceasta a fost săvârșită întrucât este rațională (capitolul 11). Pentru a face aceasta nu este destul să arătăm că are consecințe bune pentru agent: trebuie să o înțelegem ca fiind optimă din punctul de vedere al agentului. De exemplu, este un fapt acela că, dacă oamenii acordă o valoare mare unor consecințe viitoare ale unui comportament prezent, adică au o rată scăzută a deprecierii timpului (capitolul 6), atunci viața lor este mai bună. Este, de asemenea, plauzibil ca educația superioară să modeleze preferințele legate de timp într-o astfel de direcție. Totuși, aceste două premise nu sunt echivalente cu o explicație în termenii alegerii raționale a motivului pentru care oamenii iau decizia de a deveni educați. Pentru ca o explicație să fie viabilă ar trebui să arătăm că oamenii au opiniile cerute privitoare la impactul educației asupra abilității de a amâna recompensele și că sunt motivați la un nivel subiectiv să dobândească o asemenea abilitate.[1] Încercarea de a explica o alegere prin apel la consecințele benefice este o formă de „funcționalism al alegerii raționale", înseamnă, cu alte cuvinte, a combina două abordări asupra cărora l-am prevenit pe cititor încă din Introducere, ceea ce nu aduce lumină în privința semnificației comportamentului respectiv.

[1] În capitolul 11 voi argumenta că ideea de a fi motivat să fii motivat de consecințele pe termen lung este incoerentă din punct de vedere conceptual, dar acesta este un argument separat de acela pe care îl propun aici.

Dacă un comportament este rațional, atunci este *ipso facto* inteligibil. Totuși, și un comportament irațional poate fi inteligibil. Voi face o distincție între aceste varietăți de comportament inteligibil, dar irațional, și le voi pune în contrast cu cazurile de comportament ininteligibil.

Primul tip apare atunci când mecanismul din spatele luării de decizii (vezi Figura 11.1) este *trunchiat* într-un anumit mod. În virtutea urgenței sale caracteristice, o anumită emoție puternică îl poate împiedica pe agent să „se uite în jur" (adică să culeagă informații) înainte de a acționa. În locul adoptării unei strategii expectative, similară cu aceea care l-a făcut faimos pe generalul roman Fabius Cunctator („cel care ezită"), agentul se grăbește să acționeze fără să se mai gândească la consecințe. Un alt mod de trunchiere apare în cazul slăbiciunii voinței, înțeleasă în mod tradițional ca fapt de a acționa împotriva a ceea ce agentul crede că e cel mai bine (capitolul 6). Cineva care s-a decis să se lase de fumat și acceptă totuși o țigară o face dintr-un anumit motiv, acela de a-și satisface dorința de a fuma. Dar, pentru ca o acțiune să fie rațională, ea trebuie să fie optimă în lumina totalității motivelor, nu doar a unuia dintre ele. Voi discuta, însă, cu altă ocazie chestiunea înțelegerii slăbiciunii voinței.

Întâlnim al doilea tip de comportament în cazul scurtcircuitării mașinăriei de luat decizii când formarea de opinii este influențată de dorințele agentului. Gândirea deziderativă, de exemplu, este irațională, dar pe deplin inteligibilă. O formă mai subtilă de formare motivată a unei opinii apare când agentul încetează să mai culeagă informații atunci când informațiile culese până la un anumit moment susțin o opinie pe care acesta ar dori să o considere adevărată.[1] Aceste forme motivate de formare a unei opinii sunt, în felul lor, procese de optimizare: maximizează plăcerea pe care un agent o dobândește din opiniile sale despre lume, mai degrabă decât plăcerea pe care acesta o așteaptă din confruntarea efectivă cu lumea.

Al treilea tip de comportament ar putea fi numit *încălcirea firelor* în mașinăria de luat decizii. Putem înțelege cu ușurință de ce mintea se poate angaja în reducerea disonanței cognitive (un tip al acestei disonanțe fiind gândirea deziderativă), dar este mult mai dificil să înțelegem de ce mintea s-ar angaja activ în *producerea* disonanței. Ideea aceasta, pe care am menționat-o și în capitolul 2, anume că putem crede cu ușurință ceva de care ne temem, este un astfel de exemplu. De ce frica de un anumit rezultat ne poate face să îl vedem ca mult mai probabil decât o arată faptele? De ce adoptăm o opinie cu toate că nu avem dovezile care să ne facă să o credem și nici dorințele noastre nu ne împing în acea direcție? În mod evident, nu avem de-a face cu un proces de optimizare. Într-un sens, un astfel de comportament este mult mai greu de înțeles decât acțiunile care își au originea în trunchiere și scurtcircuitare, din moment ce agentul *nu*

[1] Într-o lume a inocenței statistice, Gregor Mendel, cel care a descoperit legile eredității, se pare că practica această metodă a „renunțării când ești în câștig" în cadrul experimentelor sale.

trage niciun folos din aceasta, nici măcar parțial sau pe termen scurt. Și totuși, chiar și acest comportament este inteligibil (cel puțin așa cum înțeleg eu ideea) întrucât emerge din sistemul credință-dorință al agentului.

Printre acțiunile care evită orice interpretare sunt incluse acelea cauzate de dorințele nestăpânite și de obsesii, comportamentele guvernate de fobii, automutilările, anorexia și așa mai departe. Categoric, acest tip de comportament este explicabil prin aceea că are efectul calmării anxietății, pe care agentul o simte dacă nu manifestă respectivul comportament. Totuși, a te spăla pe mâini de cincizeci de ori pe zi sau a urca cincizeci de etaje de frica liftului nu este același lucru cu a folosi un calmant. A lua valium poate fi la fel de rațional cu a lua aspirină, dar comportamentul dominat de emoții nestăpânite sau cel fobic este ininteligibil întrucât nu este parte a unui *sistem* interconectat de credințe și dorințe. Putem lua un exemplu din John Rawls: ne-ar fi greu să înțelegem comportamentul cuiva care și-ar dedica timpul numărării firelor de iarbă dacă respectiva ocupație nu ar avea legătură cu alt scop, cum ar fi câștigarea unui pariu.

Gândirea dezideravă este inteligibilă, la fel cum este și gândirea antidezideravă. Credința unui individ tulburat mintal că dentistul din clădirea vecină îl bombardează cu raze X pentru a-i distruge mintea nu este inteligibilă. Prin contrast, credințele paranoice în politică sunt inteligibile pentru că își au rădăcinile în dorințele agentului. O persoană puternic antisemită are motivația să nutrească opinii absurde despre caracterul omnipotent și malefic al evreilor (vezi capitolul 7). Persoana nu vrea ca evreii chiar *să aibă* aceste trăsături, ci este motivată să creadă că aceștia le au întrucât, astfel, își poate raționaliza imboldul de a-i distruge. Până și opiniile contradictorii pot fi inteligibile. Un antisemit poate uneori să-i caracterizeze pe evrei ca „paraziți" și, în același timp, poate să susțină că aceștia sunt omnipotenți. Aceiași oameni care susțin că „evreii încearcă întotdeauna să se bage acolo unde nu sunt doriți" crede în același timp că „evreii au trăsături tribale, stau tot timpul împreună". Unul și același musulman poate să afirme că serviciul secret israelian Mossad s-a aflat în spatele atacului de la World Trade Center din 11 septembrie 2001 *și*, în același timp, să fie mândru că evenimentul s-a întâmplat.

Înțelegerea războaielor civile

Voi da două exemple pe care le voi analiza în detaliu pentru a ilustra noțiunile de credință inteligibilă și dorință inteligibilă, ambele preluate din studii asupra războaielor civile din trecut și din prezent. Mă voi baza apoi pe aceleași studii și pe alte câteva pentru a reflecta asupra întrebării hermeneutice fundamentale referitoare la modul în care putem atribui sau stabili motivații și credințe.

Pentru început, să luăm în considerare credința în predestinare, subiect principal de discordie între calviniști și catolici în timpul războaielor religioase. La origine se afla anxietatea religioasă intensă pe care o trăiau mulți credincioși în vremea de dinaintea

Reformei, cauzată de incertitudinea privitoare la propria mântuire. Cum poate cineva să fie sigur – este aceasta, oare, o posibilitate – că a făcut tot ce trebuie pentru a o obține? Gândindu-se la anii tinereții, Calvin scria în 1539 că, până și atunci când satisfăcea toate cerințele impuse de Biserică și își mărturisea păcatele ștergând amintirea lor de către Dumnezeu prin fapte bune și penitență, „eram departe de a fi sigur și nu aveam conștiința împăcată. De fiecare dată când mă retrăgeam în mine însumi și îmi înălțam sufletul către Tine, eram cuprins de o groază atât de cumplită, că nicio formă de ispășire și nicio iertare a păcatelor nu mă putea vindeca."

Anxietatea i-a fost domolită prin mutarea accentului de pe o concepție asupra unui Dumnezeu imanent lumii, o prezență tiranică și amenințătoare, pe o concepție asupra unui Dumnezeu transcendent în mod absolut. Această idee a fost legată în mod crucial de doctrina dublei predestinări: din moment ce Dumnezeu a ales dintru eternitate cine va fi mântuit și cine va fi osândit, nimeni *nu poate face nimic* pentru mântuire și, prin urmare, preocuparea că nu ai făcut destul nu are niciun temei. Chestiunea interpretativă-cheie privește legătura dintre această credință în predestinare și alinarea anxietății. *A priori*, un astfel de efect al doctrinei pare ininteligibil. Conform învățăturilor lui Calvin, cei aleși sunt o minoritate restrânsă, care variază (în diverse enunțări ale doctrinei) de la unul dintr-o sută la unul din cinci. Ce ar putea genera mai multă anxietate decât credința că este foarte probabil să te numeri printre cei damnați și că nu poți face absolut nimic pentru a fi salvat de la focul veșnic al Iadului? Oare convertirea de la catolicism la calvinism nu echivalează literalmente cu a sări din tigaie direct în foc?

Cred că răspunsul poate fi găsit urmându-l pe Max Weber. Dată fiind credința lor în predestinare, calviniștii nu puteau susține că un efort rațional și sistematic le va aduce mântuirea, dar puteau susține, și chiar o făceau, că le dădea *certitudinea* subiectivă a mântuirii. Calvin însuși scrie că „vocația celor aleși este asemenea unei demonstrații și mărturii a alegerii lor". Convertirea la calvinism pare să fi eliminat foarte eficient incertitudinea cu privire la mântuire. Mă voi întoarce la această formă de „gândire magică" în capitolul 7. Pentru moment, vreau să scot în evidență modul în care mecanismele gemene ale gândirii deziderative și ale gândirii magice conferă inteligibilitate credinței în predestinare.

Să luăm în considerare acum inteligibilitatea motivațiilor. De ce sunt dispuși tinerii palestinieni să își dea viața în misiuni sinucigașe? Motivația principală – de a obține sau apăra o patrie – nu este greu de înțeles.[1] Este o cauză care poate fi la fel de

[1] Susținând că aceasta este motivația lor, nu neg că ar exista și altele, ca, de exemplu, dorința de glorie și faimă postumă, beneficiile materiale care revin familiei atacatorului sinucigaș, răzbunare pentru faptul că israelienii i-au ucis un prieten sau o rudă ori presiunea socială de a fi voluntar într-o misiune. După cum subliniez în Introducerea la partea a doua, sunt sceptic cu

convingătoare precum apărarea democrației în lupta împotriva lui Hitler. Ce poate părea de neînțeles este *intensitatea* motivației. Trebuie aduși în discuție și alți factori cauzali suplimentari pentru a face chestiunea inteligibilă. Voi discuta câțiva astfel de factori plauzibili și voi conchide în favoarea unuia.

Înainte de 11 septembrie, exista o opinie răspândită că teroristul sinucigaș tipic din Orientul Mijlociu este un tânăr șomer necăsătorit, poate chiar avid după sex, pentru care o mișcare religioasă poate umple golul ocupat altminteri de familie și de locul de muncă. Apoi, peste noapte, după atacul asupra World Trade Center, experții în terorism au luat decizia să „rescrie manualul". Stereotipul a fost pus în discuție chiar mai devreme din cauza faptului că unii atacatori implicați în misiunile sinucigașe sunt femei, situație frecventă, chiar dacă fluctuantă. În a doua Intifadă, folosirea femeilor în acte de terorism, unele chiar mame sau persoane cu un nivel înalt de educație, a produs o mare surpriză.

Sărăcia și analfabetismul, factori menționați adesea, au și ei o eficacitate cauzală limitată, cel puțin ca trăsături ale atacatorului sinucigaș ca individ. În cazul teroriștilor palestinieni, venitul și educația sunt, în general, mai ridicate comparativ cu nivelul mediu al populației. Explicațiile care mizează pe sărăcie sunt, de asemenea, foarte nesatisfăcătoare întrucât nu este clar cum ar putea sărăcia să genereze motivația necesară. Conform unui punct de vedere frecvent întâlnit, câștigul obținut dintr-un atac sinucigaș trebuie cântărit în raport cu costurile pierderii vieții. Dacă nu se consideră că viața are un preț prea mare, costul e mai mic. Conform unei astfel de abordări, o viață în mizerie și sărăcie valorează atât de puțin pentru individ, încât costurile comportate de sinucidere devin neglijabile. Sunt destul de sceptic că acest argument este corect, din moment ce cred că și săracii consideră, la fel ca toată lumea, că viața merită trăită. Este o descoperire destul de încetățenită în psihologie aceea că oamenii își ajustează aspirațiile astfel încât să se potrivească cu circumstanțele în care se găsesc și astfel încât să resimtă un nivel mai mult sau mai puțin constant de mulțumire („adaptarea hedonică").

Un factor mult mai plauzibil decât privarea totală este *privarea relativă*, adică prăpastia dintre așteptări și realitate, resimțită de mulți palestinieni educați, care nu au nicio perspectivă de a-și găsi un loc de muncă decent. Mobilitatea socială descendentă poate avea un efect asemănător. Și totuși, trăsăturile cele mai relevante par a fi sentimentele permanente de *inferioritate* și *resentimentul*. Prima emoție se bazează pe o comparație între sine și ceilalți, a doua pe interacțiunea dintre sine și ceilalți. În general, emoțiile care se bazează pe interacțiune sunt mai puternice decât acelea care se bazează pe comparație. Mulți dintre aceia care au scris despre teroriștii sinucigași

---

privire la puterea motivațională a beneficiilor religioase care vin sub forma accesului privilegiat în Paradis. ~

palestinieni scot în evidență resentimentul intens cauzat de umilința zilnică pe care o resimt în interacțiunea cu forțele israeliene. În afara controalelor și verificărilor degradante la care sunt supuși palestinienii, mai este și conștiința faptului că mulți israelieni îi consideră pe arabi „leneși, lași și cruzi", după cum mi-a spus un taximetrist din Ierusalim acum vreo douăzeci de ani.

Dacă această descriere este corectă, resentimentul puternic la adresa acelora care le ocupă acum patria ne dă posibilitatea să *înțelegem* disponibilitatea de a muri a teroriștilor sinucigași palestinieni. Dorința de a lupta împotriva israelienilor își datorează intensitatea faptului că este parte a unui complex motivațional mai larg. Există, însă, și un punct de vedere alternativ. Teroriștii palestinieni se află, de obicei, sub foarte strictul control al acelora care îi folosesc și care sunt dispuși să exercite presiuni suplimentare în cazul în care motivația primară eșuează în apropierea momentului în care trebuie acționat. Un atacator potențial din Irak, care fusese prins și dezarmat pentru că era vizibil nervos, a declarat că fusese închis timp de trei zile înaintea atacului într-o cameră cu un mullah care îi vorbise despre Paradis și îl hrănise cu „o supă specială care să-l facă puternic". Starea mentală care declanșează actul detonării unei bombe poate fi, prin urmare, efemer și poate constitui mai degrabă o trăsătură episodică, și nu stabilă, a persoanei respective. Chiar dacă termeni precum „spălarea creierului" sau „hipnoza" pot fi prea tari, există destule dovezi că unii dintre teroriști se aflau într-un fel de transă cu câteva minute înainte de a muri. Atunci când, așa cum se întâmplă în astfel de cazuri, o intenție este izolată de sistemul dorință-credință general al persoanei, nu mai este posibilă nicio interpretare. Comportamentul acelora care angajează teroriști și, în general, al acelora care organizează un atentat poate fi, desigur, obiect al interpretării.

O dilemă hermeneutică

Este corect să afirmi că un comportament trebuie explicat în termeni de stări mentale antecedente – dorințe și credințe – care cauzează comportamentul respectiv, dar cum putem stabili aceste cauze anterioare? Pentru a nu cădea în circularitate, nu putem folosi comportamentul în sine drept dovadă. Trebuie să căutăm alte dovezi, cum ar fi enunțuri ale agentului cu privire la motivațiile sale, acordul dintre comportamentul său nonverbal și aceste enunțuri, motivele atribuite de alții comportamentului său și acordul dintre comportamentul non-verbal al *acestora* și ceea ce ei îi atribuie. Și totuși, cum putem exclude posibilitatea ca aceste forme de comportament verbal și non-verbal să fi fost alese înadins pentru a-i face pe cei care ascultă să creadă în mod fals că funcționează o anumită motivație? Mărturisirea și declararea unor anumite motivații pot fi motivate de ceva anume. Problema este fundamentală pentru luarea deciziilor la nivel colectiv. După cum voi susține în capitolul 25, toate metodele

disponibile pentru a consolida preferințele individuale într-o decizie socială îi stimulează pe participanți să își reprezinte greșit preferințele.

Putem lua în considerare, de exemplu, motivele liderilor și adepților în războaiele civile. Părțile au declarat, iar oponenții lor le-au atribuit unul din trei motive: religie, putere și bani. Cei care declară motive religioase sunt acuzați adesea că le folosesc pentru a ascunde adevăratele motive, politice sau pecuniare. În timpul războaielor religioase din Franța (1562-1598), părțile combatante se acuzau reciproc că folosesc religia ca pretext pentru scopuri politice sau pecuniare. Anumite motive pentru astfel de acuzații existau. Henri de Navarre (mai apoi Henric al IV-lea) s-a convertit de șase ori în timpul vieții, iar ultima convertire, cea din 1593, a fost în mare măsură suspectată de oportunism. Tatăl său, Antoine de Bourbon, susținuse răspicat că credința lui e de vânzare pentru cel care dă mai mult. Acesta o acompania pe regina regentă la liturghia catolică și pe soția sa protestantă la împărtășanie. Pe patul de moarte, a căutat să fie absolvit de păcate conform riturilor ambelor religii. Cardinalul de Châtillon, un reformator de frunte, s-a căsătorit după convertire, dar și-a păstrat atât titlul de cardinal, cât și veniturile obținute din episcopat. Un alt prelat, Antoine Carracioli, Episcop de Troyes, a vrut de asemenea să îmbine preoția protestantă cu veniturile aduse de episcopatul catolic. Un catolic însemnat, Ducele Henri de Guise, era cât se poate de dispus să stabilească o alianță cu calviniștii împotriva regelui Henric al III-lea.

Și în zilele noastre religia este adesea folosită ca pretext pentru politică, iar politica este folosită ca pretext pentru bani. Scopurile insurgenților ceceni și ale unor organizații palestiniene, dintre care putem aminti Fatah, au fost inițial exclusiv politice. Atunci când și-au asumat și caracteristici religioase, scopul a fost acela de a atrage mai mulți adepți. În Palestina, organizația nu putea supraviețui altfel, dată fiind rivalitatea cu mișcarea incontestabil religioasă Hamas. În Filipine, grupul terorist Abu Sayyaf a folosit revendicarea unui stat islamic independent ca pretext pentru a răpi oameni și a cere recompense uriașe. În Columbia, încă este neclar dacă Forțele Armate Revoluționare ale Columbiei (FARC) și-au păstrat motivația inițială, de a lupta împotriva nedreptăților sociale, sau dacă a degenerat între timp într-o mafie. În toate aceste cazuri, ca și în războaiele religioase din Franța, atribuirea de motive este de cele mai multe ori încărcată de incertitudini. Poate fi foarte dificil de aflat dacă, de pildă, motivele liderilor și ale adepților sunt în concordanță deplină.

Există multe motive pentru care oamenii ar putea dori să își reprezinte greșit motivațiile lor și ale oponenților lor. Pe de o parte, fiecare societate are o ierarhie normativă a motivațiilor (capitolul 4), care induce dorința de a te arăta însuflețit de motivații nobile, și nu de unele vulgare, atribuind în același timp o motivație de joasă speță oponentului. Atât în războaiele religioase din Franța, cât și în Războiul Civil din Anglia fiecare dintre părți se prezenta drept motivată religios și califica partea adversă drept avidă doar după putere. Pe de altă parte, dacă îi poți convinge pe alții să aibă

încredere în faptul că lucrează în conformitate cu o anumită motivație, atunci poate fi mult mai ușor să îți atingi scopurile. Întrucât imaginea unui terorist poate fi mult mai descurajantă decât aceea a unui criminal de drept comun, răpitorii mercenari își pot spori șansele de a obține anumite concesii dacă susțin că luptă pentru o anumită cauză. În Columbia există multe cazuri de răpiri comise de criminali de drept comun care încearcă să semene frica în rândul familiilor victimelor susținând că fac parte dintr-o grupare de gherilă. Răpirile sunt mult mai înfricoșătoare dacă există percepția că teroriștii sunt dispuși să ia măsuri drastice în cazul în care ceva nu merge conform planului și că sunt mult mai puțin dispuși să se târguiască pe termene-limită sau pe bani. Dacă nu pot obține ceea ce vor, cel puțin pot „să își spună punctul de vedere" ucigându-și victimele.

Problema părtinirii, care servește propriului interes la nivelul enunțurilor privitoare la intențiile agenților sociali, este una serioasă, dar nu insurmontabilă. Un mod de a o ocoli este acela de a lua în considerare *interesele obiective* ale agentului, presupunând că, în absența unei dovezi puternice care să indice contrariul, acestea coincid cu motivațiile subiective, indiferent ce susține agentul cu privire la cele din urmă. Sau se pot determina *consecințele actuale* ale acțiunilor unui agent, presupunând că, în absența unor dovezi puternice care să indice contrariul, rezultatul este cel intenționat. (Fiecare dintre aceste idei poate fi aplicată exemplului referitor la alegerea educației superioare, discutat mai devreme.) Faptul că există aceste *două* proceduri pentru a inversa sarcina probei sugerează, însă, că niciuna nu este acceptabilă. Atât interesele obiective, cât și consecințele actuale pot sugera ipoteze utile cu privire la motivele subiective, dar niciuna nu are vreun temei în plus în favoarea sa.

Istoricii și experții din științele sociale au dezvoltat alte moduri prin care putem rezolva problema și care, mai ales atunci când sunt combinate, pot duce la concluzii rezonabil de certe. O tehnică este aceea de a depăși nivelul enunțurilor publice și de a le căuta pe acelea cu cea mai mică probabilitate de a fi motivate de dorința de a fi reprezentate greșit. Scrisorile, jurnalele, conversațiile păstrate în scris, ciornele și altele de acest fel pot constitui surse neprețuite. Știm din scrisorile adresate soțiilor că unii delegați din Adunarea Franceză Constituantă din 1789 au votat împotriva bicameralismului și votului regal pentru că au crezut că, dacă votează altfel, viața le este în pericol. În Adunare și-au justificat votul prin interesul public. În încercarea de a descoperi motivațiile din spatele masacrului din Noaptea Sfântului Bartolomeu, din 1582, istoricii au considerat util să meargă dincolo de relatările părtinitoare ale participanților și s-au bazat pe istorisirile unor diplomați străini, care aveau tot interesul să înțeleagă cum stau lucrurile de fapt. În Anglia secolului al XIX-lea, mărturiile făcute pe patul de moarte erau exceptate de la regulile obișnuite, aplicate probelor bazate pe zvonuri. Prima versiune a unui document poate spune mai multe despre credințele și motivele autorului său decât versiunea publicată. Este foarte instructiv, de pildă, să

comparăm ciornele lucrării lui Marx *Războiul Civil din Franța* sau scrisoarea sa către Vera Zasulici cu versiunile oficiale.

Poate exista, de asemenea, un contrast puternic între ceea ce spun actorii în public și ceea ce spun în spatele ușilor închise. Cu toate că dezbaterile publicate ale Adunării Constituante Franceze din anii 1789-1791 sunt fascinante, există doi factori care conspiră în a le face surse nu tocmai de încredere pentru a proba stările mentale ale agenților. Pe de o parte, faptul că totul se desfășura în public îi constrângea pe delegați să folosească doar argumente care țineau de interesul public; purul interes de grup era inadmisibil. Pe de altă parte, vanitatea lor era stimulată de faptul că se adresau unui public compus dintr-o mie de colegi delegați, pe lângă încă alți o mie de spectatori în public. Din acest punct de vedere, Convenția Federală Americană era mult mai probabil să conducă la declarații sincere. Întrucât numărul delegaților era mic (55, comparativ cu 1200 în Paris), iar desfășurarea lucrărilor rămânea învăluită în mister, putea să apară târguiala în funcție de interes, și chiar așa s-a întâmplat. În același timp, după cum a spus Madison mult mai târziu, „dacă membrii Convenției s-ar fi angajat public de la început, ar fi trebuit ca ulterior să fie consecvenți și să-și mențină punctul de vedere. Dar, pentru că discuțiile s-au purtat în secret, nimeni nu s-a simțit obligat să mențină un anumit punct de vedere, decât dacă acesta era adecvat și adevărat, fiind astfel deschis la puterea argumentului." Nimeni nu trăia cu frica faptului că lucrurile dezbătute vor fi date în vileag întrucât caracterul lor secret trebuia să rămână astfel indefinit de mult. Secretul a fost rupt abia multe decenii mai târziu, odată cu publicarea notițelor lui Madison. Motivele strategice pentru reprezentarea greșită a motivațiilor nu-și mai au rostul dacă sinceritatea nu implică niciun cost.

Experții din științele sociale pot îndepărta costul implicat de sinceritate și prin crearea unui văl artificial de ignoranță. Să presupunem că un om de știință vrea să studieze relația dintre orientarea sexuală și o altă variabilă de interes. Poate fi dificil să obții de la subiecți răspunsuri sincere la întrebări legate de experiențele lor sexuale cu persoane de același sex, chiar și sub protecția anonimatului. Pentru a evita problema, cercetătorul poate să le dea subiecților instrucțiunea de a răspunde sincer în cazul în care au avut o astfel de experiență și, în cazul în care nu au avut nicio experiență de acest tip, să răspundă da sau nu, în funcție de rezultatul aruncării cu o monedă. Dacă subiecții se supun – și nu au niciun motiv să nu o facă – și dacă eșantionul ales este suficient de mare, datele vor fi la fel de bune ca și cum toată lumea ar fi răspuns sincer.

O altă tehnică pentru a verifica dacă comportamentul non-verbal al agenților se află în acord cu motivația declarată este ca aceștia să fie întrebați: cât de mult sunteți dispuși să sacrificați în sprijinul celor afirmate? Când, în 2003, administrația Bush a susținut că motivul principal pentru invazia Irakului era certitudinea faptului că Saddam Hussein avea arme de distrugere în masă, a luat oare această administrație toate măsurile necesare pentru a-și proteja propriii soldați de amenințarea

respectivelor arme? Anumite trăsături comportamentale pot scoate la iveală motivația adevărată a unei răpiri. În 1996, în Costa Rica, niște răpitori (în mare parte, foști rebeli din Nicaragua) au cerut un milion de dolari răscumpărare, pe lângă anumite garanții legate de locurile de muncă, o reducere a prețurilor la alimente, creșterea salariului minim pe economie și eliberarea unor tovarăși rebeli din închisoare. După ce li s-au oferit 200 000 de dolari, s-au declarat mulțumiți și nu au mai insistat asupra doleanțelor de natură politică, fapt care a convins autoritățile că atitudinea de rebel/Robin Hood era doar un truc și că, de fapt, banii fuseseră scopul principal de la bun început. Putem să ne referim și la atitudinea aristocraților francezi emigrați în Londra în timpul Revoluției Franceze. În plin focar al zvonurilor legate de o restaurație iminentă a monarhiei și în mijlocul unei competiții a declarării adeziunii față de monarhie, era vital să comunici disponibilitatea de a servi contrarevoluția. Asigurările prin viu grai nu erau însă de ajuns. Orice persoană care lua în chirie un apartament pentru mai mult de o lună era privită cu neîncredere; cel mai bine era să închiriezi cu săptămâna pentru a da impresia disponibilității de a fi rechemat în Franța de mișcarea contrarevoluționară.

Nu doar contemporanii, ci și istoricii se folosesc adesea de astfel de indicatori comportamentali pentru a judeca sinceritatea declarațiilor publice de loialitate. Spre sfârșitul celui de-al Doilea Război Mondial, de exemplu, exista un grad sporit de scepticism în Franța aflată sub ocupație cu privire la șansele de izbândă în război ale Germaniei. Nu era deloc sigur să exprimi public o astfel de atitudine, dar ea se reflecta în comportament. Procentul elevilor de liceu (sau al părinților) care alegeau germana ca limbă străină s-a dublat din 1939 până în 1942, pentru a scădea cu rapiditate după această dată. Multe edituri care, până atunci, s-au grăbit să semneze pentru drepturile de a traduce cărți germane au ales să nu se folosească de această opțiune.

Judecătorii și juriații urmează de multe ori aceeași cale. Uneori aceștia întreabă: „A avut acuzatul vreun motiv să comită fapta X?", sperând că un răspuns îi va ajuta să decidă dacă persoana a comis într-adevăr fapta X. În acest caz, „a avea un motiv" este o idee obiectivă, și anume dacă acuzatul ar avea ceva de câștigat de pe urma faptei X. În alte cazuri, mai relevante pentru ideea de față, s-a stabilit deja că acuzatul a comis X, iar întrebarea este: „Care a fost motivul pentru a comite X?" Pentru a stabili dacă o crimă a fost pasională sau o acțiune comisă cu sânge-rece, judecătorii și juriații nu investighează doar beneficiile obiective, ci încearcă să stabilească starea mentală subiectivă a acuzatului. Dacă acuzatul susține că a acționat într-un moment de mânie sau gelozie și apoi se dovedește că a cumpărat arma crimei din timp sau că a comis crima într-un timp mai îndelungat[1], atunci credibilitatea acuzatului scade.

[1] Într-o decizie judecătorească din Marea Britanie (cazul R. *vs.* McPherson) din 1957, Lordul Goddard se întreabă retoric: „Cum se poate susține că acuzatul a acționat într-un

Luate individual, oricare dintre aceste tehnici poate eșua. Un agent de poliție poate să nu fie dispus să recunoască față de soția sa că s-a temut pentru viața lui sau poate susține că i-a fost frică pentru a ascunde un motiv mai puțin respectabil (de exemplu, că a luat mită). În India secolului al XIX-lea, mărturiile făcute pe patul de moarte nu erau considerate de încredere întrucât, de multe ori, se întâmplă ca oamenii să-și folosească ultimul cuvânt pentru a le face rău dușmanilor lor. În exemplul de mai sus, cu emigrații francezi, atât adepții adevărați ai monarhiei, cât și aceia care nu credeau că monarhia se va mai întoarce puteau fi motivați să închirieze cu săptămâna, cei dintâi pentru a facilita întoarcerea în Franța atunci când va fi cazul, iar cei din urmă pentru a nu părea defetiști. Există limite, totuși, ale abilității oamenilor de a țese pânza încurcată a înșelăciunii fără a-și da în vileag motivele adevărate. Ipocrizia, după cum spune Somerset Maugham, este o activitate care îți ocupă tot timpul. Chiar și Tartuffe s-a încurcat la final. Pentru a argumenta în favoarea convingerilor religioase ale lui Henric al IV-lea, biograful său nu citează doar dovezile pozitive ale unor „numeroase episoade în care spiritul său religios s-a manifestat fără nicio intenție de afișare", dar susține și că, „dacă ar fi existat urmă de ipocrizie, aceasta și-ar fi arătat chipul în cutare sau cutare moment agreabil". În același spirit îl putem cita pe Montaigne:

> Aceia care îmi contrazic declarațiile spunând că francheța, simplitatea și naturaleța mea sunt simple artificii și prudență vicleană, mai degrabă decât bunătate, forțate, mai degrabă decât firești, mânate de bun-simț, mai degrabă decât de noroc, mai mult îmi fac cinste decât necinste. Cu siguranță, ei fac viclenia mea să pară prea vicleană. Dacă oricare dintre aceștia m-ar urmări îndeaproape și m-ar spiona, l-aș declara câștigător, dacă nu ar admite că nu e nicio învățătură propovăduită de secta lui care să falsifice modul meu firesc de a lucra și să mențină aparența unei libertăți cumpănite traversând căi atât de încâlcite, și nici nu ar putea să susțină o libertate de acțiune atât de lipsită de compromisuri pe căi atât de diferite fără să admită că toată strădania și istețimea lor nu i-ar îndemna să acționeze în același mod.

Cu toate că beneficiile reprezentării greșite pot fi însemnate, costurile pot fi prohibitive. Până la un punct, declararea de motive cu anumite scopuri este autolimitantă. Întrucât orice motiv este inclus într-o rețea vastă de alte motive și credințe, numărul de ajustări care trebuie făcut pentru a susține ipocrizia poate crea un handicap. O singură notă falsă poate fi suficientă pentru a prăbuși întreaga construcție. Multe proverbe stau mărturie pentru faptul că încrederea, odată trădată, este imposibil de recuperat. Deși credința populară conform căreia „Cine spune o minciună va spune o

---

moment de pasiune când a tras nu unul, ci patru gloanțe, fiecare presupunând eliminarea cartușelor și reîncărcarea armei?"

sută" are nevoie de o contextualizare serioasă (capitolul 10), credința recontextualizată este întâlnită destul de des și folosește în bună măsură pentru a descuraja minciuna. Din acest motiv poate că Descartes a avut dreptate, printre altele, în a spune: „Cea mai mare subtilitate dintre toate este aceea de a nu te folosi niciodată de subtilitate."

◈◈◈

Notă bibliografică

Pentru o amplă prezentare a dezbaterii legată de *Erklären* versus *Verstehen* se poate consulta partea a III-a a lucrării lui M. Martin și L. McIntyre (editori), *Readings in the Philosophy of Social Science* (Cambridge, MA, MIT Press, 1994). Capitolul lui Dagfinn Føllesdal din acel volum, „Hermeneutics and the hypothetico-deductive method", argumentează în favoarea unei poziții foarte apropiate de a mea. Citatul din Weber se găsește în eseul său „The interpretive understanding of social action", în M. Brodbeck (ed.), *Readings in the Philosophy of the Social Sciences* (Londra, Macmillan, 1969), p. 33. Incoerența atitudinilor antisemite este abordată în J. Telushkin, *Jewish Humor* (New York, Morrow, 1992), care este sursa mea și pentru alte observații cu privire la și aparținând unor evrei despre presupusele lor caracteristici. Observațiile legate de alegerea educației sunt o polemică implicită împotriva lucrării lui G. Becker și C. Mulligan, „The endogenous determination of time preferences", *Quarterly Journal of Economics*, 112 (1997), pp. 729-758. Valoarea de probă în instanță a confesiunilor de pe patul de moarte este discutată în J. F. Stephen, *A History of English Criminal Law* (Londra, Macmillan, 1883; Buffalo, NY, Hein, 1964), vol. 1, pp. 447-449. H. Sass, „Affektdelikte", *Nervenarzt*, 54 (1983), pp. 557-572, enumeră treisprezece motive pentru care afirmația de a fi comis o crimă pasională poate să nu fie credibilă. O discuție interpretativă excepțională a motivațiilor din timpul războaielor religioase din Franța se găsește în D. Crouzet, *Les guerriers de Dieu* (Paris, Camp Vallon, 1990). Analize interpretative ale motivațiilor și credințelor teroriștilor sinucigași se găsesc în eseurile scrise de S. Holmes, L. Ricolfi și de mine, în D. Gambetta (ed.), *Making Sense of Suicide Missions* (Oxford University Press, 2005). Comentariul cu privire la credințele religioase ale lui Henric al IV-lea se găsește în J.-P. Babelon, *Henri IV* (Paris, Fayard, 1982), p. 554. Scepticismul extrem cu privire la motive este discutat în G. Mackie, „Are all men liars?", în J. Elster (ed.), *Deliberative Democracy* (Cambridge University Press, 1998).

II
GÂNDIREA

Cartea de față este organizată în jurul unui model al acțiunii, și anume „modelul credință-dorință". Pentru a înțelege modul în care oamenii acționează și interacționează trebuie să fi înțeles mai întâi modul în care funcționează mintea lor, chestiune care ține, în mare parte, de introspecție și de psihologia populară, rafinată și corectată de studiile mai sistematice întreprinse de psihologi și, din ce în ce mai mult, de aceia care cercetează economia comportamentală. Modelul este vital nu doar pentru explicarea comportamentului, dar și pentru atribuirea de laude, vinovăție sau pedeapsă. De obicei, vina presupune *mens rea*, intenții și credințe. Răspunderea strictă – vina atribuită doar pe baza consecințelor realizate ale unei acțiuni – este rară. De fapt, uneori atribuim vina chiar atunci când există doar intenția, fără să se fi materializat vreo consecință. Tentativa de omor este o crimă. Conform lui John Donne, „vrăjitoarele se gândesc uneori că ucid atunci când nu o fac, prin urmare, sunt la fel de vinovate ca și cum ar fi făcut-o". Hobbes scrie: „În ceea ce le privește pe vrăjitoare, nu cred că vrăjitoria lor este o putere reală, dar cred că sunt pe drept pedepsite pentru credința falsă că ar fi capabile să provoace daune, dimpreună cu intenția de a le provoca dacă sunt în măsură."

Modelul credință-dorință, deși indispensabil, este fragil. Metodele pe care le folosim pentru a atribui stări mentale altor oameni nu produc întotdeauna rezultate stabile. Dacă vrem să măsurăm înălțimea unei clădiri, nu are importanță dacă măsurăm de la acoperiș spre parter sau de la parter în sus. În determinarea credințelor și dorințelor rezultatul poate depinde de factori irelevanți. Luați în considerare, de exemplu, ideea că oamenii „maximizează utilitatea așteptată" (capitolul 11). Mai exact, trebuie să presupunem că aceștia au o idee clară și stabilă privitoare la *valoarea* acordată fiecărui rezultat posibil al unei acțiuni, și la *probabilitatea* pe care o atribuie apariției respectivului rezultat. De multe ori, o astfel de presupoziție este justificată, dar alteori nu.

Să luăm în considerare credințele agentului. Când vrem să extragem probabilitățile subiective pe care o persoană le acordă unui anumit eveniment, o procedură standard este următoarea. Începând cu un număr p, întrebăm acea persoană dacă ar

prefera o loterie în care câștigă o anumită sumă de bani cu probabilitate p sau una în care câștigă aceeași sumă de bani dacă evenimentul în cauză se întâmplă.[1] Dacă preferă prima opțiune, expunem persoana unei noi alegeri cu probabilitatea ajustată în scădere; dacă preferă a doua opțiune, ajustăm probabilitatea în creștere. Urmând în continuare această procedură, vom ajunge în cele din urmă la o valoare de probabilitate p^*, astfel încât pentru acea persoană nu mai contează diferența dintre o loterie în care câștigă bani cu probabilitatea p^* și una în care câștigă dacă evenimentul se întâmplă. Putem afirma apoi că probabilitatea relevată sau extrasă pe care o atașează evenimentului este p^*. În principiu, p^* ar trebui să fie independent de valoarea inițială a lui p: adică, probabilitatea extrasă ar trebui să fie independentă de procedura de extragere. În practică, lucrurile nu stau chiar așa: o valoare mai mare a lui p determină o valoare ridicată a lui p^*. Această descoperire sugerează că, până la un punct, cel puțin, *nu există nicio stare de fapt*, nicio stare mentală stabilă care să fie surprinsă de această procedură.[2]

Alte proceduri sunt încă și mai fragile. De multe ori, cercetătorii atribuie probabilități subiective agenților bazându-se pe presupoziția că, atunci când aceștia au un nivel limitat de cunoaștere a situației, vor atribui valori egale de probabilitate pentru fiecare dintre stările posibile ale lumii. Justificarea unei astfel de proceduri se presupune a fi „principiul rațiunii insuficiente": dacă nu ai motive temeinice pentru a te gândi că o stare a lumii este mai probabilă decât alta, logica te forțează să atribui valori de probabilitate egale pentru fiecare dintre ele. Dar stările lumii pot fi conceptualizate și numărate în moduri diverse. Să presupunem că ești în urmărirea unui hoț și ajungi la o răscruce, unde drumul se desparte în trei, iar două căi duc la deal și alta la vale. Din moment ce nu ai niciun motiv pentru a te gândi că hoțul a pornit-o pe vreuna dintre cele trei căi anume, probabilitatea ca el să o fi luat la vale ar trebui, conform principiului, să fie egală cu o treime. Dar, pentru că nu ai niciun motiv să crezi că a luat-o la deal mai degrabă decât la vale, aceeași probabilitate ar trebui să fie 1/2. Cel puțin în acest caz, principiul rațiunii insuficiente este prea nedeterminat pentru a ne fi de folos în construcția sau atribuirea de valori de probabilitate.

---

[1] Trebuie să presupunem că evenimentul în cauză este unul care, dacă s-ar fi întâmplat, nu ar fi afectat persoana respectivă în vreun fel, ca în cazul descoperirii vieții pe alte planete. Dacă evenimentul considerat este victoria echipei sportive preferate, atunci persoana ar putea paria bani că echipa va pierde, astfel încât, indiferent de ceea ce se întâmplă, ceva tot îi va produce mulțumire.

[2] Probabil că acest enunț este prea tare. Manipularea procedurii ar putea duce la extragerea oricărei valori de probabilitate între 50% și 80%, dar niciuna în afara acestui interval. Într-un astfel de caz, ar fi justificat să spunem că subiectul crede că evenimentul este mult mai probabil să se întâmple decât să nu se întâmple, dar nu este sigur că se va întâmpla. Această evaluare este, cu toate acestea, mult mai grosieră decât ar fi necesar pentru modelele standard ale luării de decizii.

Să luăm în considerare, în cele ce urmează, extragerea preferințelor. Au existat experimente în care subiecții au fost întrebați dacă ar cumpăra anumite obiecte (accesorii de calculator, sticle de vin etc.) la un preț egal cu ultimele două cifre ale codului numeric personal. Așadar, subiecții au fost rugați să enunțe prețul maxim pe care ar fi dispuși să îl plătească pentru produs. Codul numeric personal a fost evidențiat ca având un impact semnificativ asupra obiectelor pentru care sunt dispuși să plătească. De exemplu, subiecții al căror cod numeric personal se încadra în prima cincime erau dispuși să plătească în medie 56 de dolari pentru o tastatură fără fir, pe când cei din ultima cincime erau dispuși să plătească 16 dolari. Deși procedurile fuseseră gândite pentru a pune în evidență sau pentru a extrage preferințe preexistente, rezultatele au indicat că nu exista nimic care să fie extras, nicio stare de fapt. Numerele datorau mai mult faptului că făceau parte din codul numeric personal, și nu unor preferințe „reale".

Există, de asemenea, probe care arată că și *compensațiile* între valori sunt foarte instabile și pot datora mai mult artificiilor procedurale decât unei realități mentale subiacente. Compensațiile pot fi conceptualizate în cadrul experimentelor sau prin ideea de *alegere* ori prin aceea de *potrivire*. Subiecților li se dă posibilitatea să aleagă între a salva un număr mare de vieți la un cost mare per viață salvată (A) sau aceea de a salva un număr mai mic de vieți la un cost mai mic (B). Alternativ, li se poate cere să indice costul per viață salvată care i-ar face indiferenți între a salva numărul mai mare de vieți la respectivul cost (opțiunea C) și opțiunea B. Să presupunem că un anumit subiect enunță un cost mai mic decât cel al lui A. Din moment ce persoana este indiferentă între C și B și putem presupune că preferă C lui A (întrucât C salvează mai multe vieți la un cost mai mic), ar trebui ca persoana să prefere și, prin urmare, să aleagă B în loc de A. Majoritatea covârșitoare a subiecților au ales, într-adevăr, un cost pentru C sub cel al lui A; totuși, două treimi dintre subiecți au afirmat că ar alege A în loc de B. Valoarea mai importantă – cea a salvării de vieți – este mult mai puternică în alegere decât în potrivire, deși cele două proceduri ar trebui să fie echivalente.

Există și alte motive pentru care nu ar trebui să tratăm enunțurile despre credințe și alte stări mentale conform valorii aparente. Credințele religioase sunt în special problematice din acest punct de vedere. În Anglia secolului al XVII-lea timpuriu, un prelat precum Episcopul Andrewes putea să susțină că ciuma este o pedeapsă a lui Dumnezeu asupra păcătoșilor și, în același timp, să părăsească Londra pentru a merge la țară. Credința că, în virtutea originii lor divine, regii Franței pot vindeca scrofuloza doar prin atingerea bolnavului își pierduse mult din tărie la sfârșitul secolului al XVIII-lea, când formularea tradițională („Dacă regele te atinge, Dumnezeu te vindecă") a fost înlocuită de o nouă versiune („Dacă regele te atinge, fie ca Dumnezeu să te vindece"). Înflăcărarea cu care curtenii căutau să documenteze cu probe vindecările care au avut succes sugerează faptul că avem de a face cu o credință care nu era prea sigură pe sine.

Iată un exemplu mai aproapiat de zilele noastre: comportamentul teroriștilor sinucigași islamici poate fi explicat, cel puțin în parte, de credința lor într-o viață de apoi la care le dă acces privilegiat statutul de martir. Ne putem întreba dacă această „credință" este de aceeași natură ca și credința noastră că soarele va răsări din nou mâine, altfel spus, dacă este utilizată cu aceeași încredere drept premisă pentru acțiune. Aceasta nu este o chestiune legată de certitudine *versus* probabilitate, ci de încredere *versus* lipsa ei. Pot avea mare încredere într-o credință care se bazează pe multe întâmplări trecute – și chiar sunt dispus să pariez pe așa ceva. Credința în viața de apoi nu are, probabil, un astfel de caracter. Mai degrabă, este un fel de „cvasi-credință", destul de tenebroasă, care este întreținută pentru valoarea sa de consum, și nu ca o premisă a acțiunii. Dacă toți aceia care susțin că cred în viața de apoi ar întreține această credință cu o certitudine deplină sau cel puțin cu o „probabilitate care asigură încrederea", atunci am observa mult mai mulți martiri decât vedem deja. Deși unii credincioși pot fi de acest tip, iar teroriștii sinucigași pot fi recrutați într-un mod disproporționat din acest subset, am impresia că, pentru cei mai mulți, religia servește drept consolare odată ce decizia de a merge într-o misiune a fost deja luată și nu constituie o premisă a acțiunii.[1]

Similar, oamenii pot experimenta sau pot susține că experimentează „cvasi-emoții", care diferă de emoțiile autentice prin faptul că nu au niciun amestec în acțiune. Unii oameni care susțin că sunt indignați de sărăcia din țările din Lumea a Treia, cu toate că nu pun mâna niciodată pe portofel pentru a face ceva în această privință, se pot bucura de indignarea respectivă ca de un bun de consum, întrucât acest lucru îi face să aibă o părere bună despre sine. Asemănător, bucuria vizibilă a multora dintre cei care au susținut că sunt îndurerați (sau „cvasi-îndurerați") de moartea prințesei Diana era incongruentă cu sentimentul oribil al durerii autentice. Termenul adecvat pentru astfel de sentimente cred că este „sentimentalism". (Termenul german *Schwärmerei* este mult mai potrivit.) Oscar Wilde a definit omul sentimental drept „o persoană care dorește să se bucure de luxul unei emoții fără să plătească pentru ea". Indiferent dacă plata ia forma unor donații pentru Oxfam sau forma suferinței, putem să ne dăm seama, din absența ei, că nu avem de a face cu emoția reală.[2]

O problemă înrudită este imensa putere a autosugestiei. Odată ce știm că un X este presupus a fi un Y, susținem și chiar *credem* că esti în mod evident un Y. Cei mai

---

[1] Ideea că religia este „opiumul popoarelor" sugerează, de asemenea, că este un bun de consum, și nu o premisă a acțiunii. Religia poate constitui, însă, o premisă a inacțiunii.

[2] Un fenomen înrudit apare atunci când oamenii adoptă perspectiva celei de-a treia persoane asupra lor înșiși. Un proverb spune: „Virtutea nu se cunoaște pe sine." De asemenea, nu poți să îți afirmi propria naivitate, din moment ce ideea în sine presupune exact lipsa conștiinței de sine. După cum o spune Nero Wolfe, într-unul dintre romanele lui Rex Stout în care apare, „a afirma demnitatea înseamnă a renunța la ea".

mari experți ai lumii în pictura lui Vermeer au fost păcăliți de (ceea ce *acum* par a fi) evidente tablouri contrafăcute ale lui van Meegeren. Proust vorbește despre „acea aptitudine care îți dă posibilitatea să descoperi intențiile unei bucăți simfonice atunci când i-ai citit programul sau trăsăturile familiare ale unui copil căruia îi cunoști rudele apropiate". Un fan european al jazzului și-a schimbat în totalitate aprecierea pentru Jack Teagarden când a aflat că acesta nu e negru. Dacă manifestăm preferințe pentru un anumit scriitor, putem citi semnificații profunde în observații pe care un ochi neutru le-ar considera banale. Ne proiectăm așteptările asupra lumii și apoi pretindem ca lumea să ne confirme și să ne justifice credințele.

O consecință a acestor observații este aceea că ar trebui să fim precauți în a ne gândi la credințe, dorințe, preferințe, emoții etc. ca fiind entități stabile și durabile, similare cu obiecte precum merele și planetele. Voi furniza în capitolele următoare mai multe exemple care să ilustreze natura evazivă, instabilă sau dependentă de context a stărilor mentale. Acestea fiind zise, enunțurile care par să exemplifice tipul de pseudo-empirism sau rigoare simulată, asupra cărora am atras atenția, pot fi descoperite și de cititori. Pentru a merge mai departe de enunțurile calitative, care nu au semnificații clare, ca, de exemplu, „Agenții acordă o greutate mai mare bunăstării prezente decât celei viitoare", trebuie să putem fi în măsură să spunem *cât de multă* greutate acordă ei prezentului. Odată ce am făcut aceasta, enunțurile noastre vor conține inevitabil distincții mult mai precise decât ceea ce se poate citi în comportamentul agenților. Artificiul – mai mult o artă decât o știință – este a ști când simplificările pun în lumină și când distorsionează. În cartea de față mă voi referi în repetate rânduri la activitățile necunoscute ale minții. De exemplu, reducerea disonanței (capitolul 1), gândirea dezideravă (capitolul 7) și transmutația motivelor (capitolul 4), sunt cauzate de mecanisme inconștiente. Deși nu înțelegem prea bine modul în care funcționează acestea, este imposibil să negăm faptul că ele există. Mulți au susținut chiar că există *stări* mentale inconștiente. Autoînșelarea, spre deosebire de gândirea dezideravă, presupune existența unor credințe inconștiente. Freud credea că toți avem anumite dorințe inconștiente și neconfesabile. Ar putea exista chiar emoții și prejudecăți inconștiente.

În măsura în care stările mentale au eficiență cauzală, ar trebui să fie posibilă identificarea lor după efectele produse. Dacă, de exemplu, negarea unui enunț este disproporționat de puternică, putem infera că persoana respectivă crede respectivul enunț, cu toate că o face inconștient: „Cred că regina face prea multe jurăminte."[1] Există o anecdotă (căreia nu am reușit să îi găsesc sursa) despre Sigmund Freud, care a fost invitat să se întâlnească cu o persoană de renume, Dr. X, în cadrul mișcării evreiești internaționale. În timpul conversației, Dr. X îl întreabă: „Spuneți-mi, Dr. Freud, cine este, în opinia dumneavoastră, cea mai importantă personalitate de origine ebraică din lume la momentul actual? Freud a răspuns politicos: «Păi, cred că dumneavoastră

[1] Shakespeare, *Hamlet*, Actul III, Scena II. (N. tr.)

sunteți aceea, Dr. X.» Dar când Dr. X i-a replicat: «Nu, nu», Freud l-a întreabat: „Oare nu ar fi fost suficient un singur *nu*?" Dubla negație poate echivala cu o afirmație.

De asemenea, putem identifica prejudecățile inconștiente prin efectele lor. În cadrul unor experimente, subiecților li s-a cerut mai întâi să clasifice foarte rapid (prin simpla atingere a genunchiului stâng sau drept) o serie de nume în două categorii: în prima, numele considerate a fi asociate de obicei unor persoane de culoare (cum ar fi Malik sau Lashonda) și, în a doua, numele asociate de obicei albilor (cum ar fi Tiffany sau Peter). Apoi li s-a cerut să clasifice foarte rapid mai multe cuvinte după cât de plăcute sunt ca înțeles (ca, de exemplu, „iubire" și „bebeluș") sau după cât de neplăcute sunt (de exemplu, „război" și „vomă"). Apoi au alcătuit o o listă de cuvinte ordonate întâmplător, care includea toate numele de persoane de culoare, toate numele albilor, cuvintele plăcute și neplăcute. Mai întâi li s-a cerut să își atingă genunchiul stâng pentru fiecare nume de persoană de culoare și pentru fiecare cuvânt cu înțeles neplăcut și genunchiul drept pentru fiecare nume de alb și cuvânt plăcut. Apoi instrucțiunile au fost schimbate. Au fost rugați să își atingă genunchiul stâng pentru numele de albi și pentru cuvintele neplăcute, iar genunchiul drept pentru numele persoanelor de culoare și pentru cuvintele plăcute. A durat de două ori mai mult pentru a îndeplini cea de-a doua sarcină, deși, din punct de vedere obiectiv, cele două sarcini erau de dificultate egală.

Emoțiile inconștiente pot fi adesea identificate de observatori, care le deduc existența din expresiile fiziologice sau comportamentale caracteristice. Mulți dintre noi am auzit și am rostit la nervi enunțul „Nu sunt nervos". Invidia se poate manifesta printr-un ton ascuțit sau printr-o tendință de a adopta o atitudine peiorativă, care sunt evidente pentru observatorii neutri, dar nu și pentru subiect. În *Roșu și negru*, Doamna de Rênal își descoperă sentimentele pentru Julien Sorel doar când suspectează că acesta ar putea fi îndrăgostit de cameristă, o emoție (gelozia) dezvăluind prezența alteia (iubirea).

Autoînșelarea (vezi capitolul 7) este mai problematică din acest punct de vedere. Să presupunem că îmi formez, apoi îmi reprim convingerea că soția mea mă înșală cu cel mai bun prieten. Deși inconștientă, credința că aceștia sunt iubiți îmi poate călăuzi acțiunile; de exemplu, mă poate împiedica să merg în acea parte a orașului unde locuiește prietenul meu ca să nu risc să o surprind pe soția mea în timpul unei vizite. Poate că aceasta este o istorisire plauzibilă, dar, din câte știu, nu există vreo dovadă care să indice că credințele inconștiente au eficiență cauzală. Multe argumente în favoarea existenței autoînșelării se bazează pe (1) expunerea persoanei la dovezi puternice care să sugereze o credință despre care aceasta nu ar dori să fie adevărată și (2) pe faptul că agentul întreține o credință diferită mult mai acceptabilă pe baza căreia acționează. Pentru a obține o dovadă directă pentru persistența inconștientă a credinței

inacceptabile ar trebui să arătăm (3) că și aceasta este capabilă să călăuzească acțiunea, ca în exemplul ipotetic de mai sus. Repet, nu cunosc vreo demonstrație în acest sens.

Bănuiala mea este că fenomenul nu există. Ar fi frumos să ne imaginăm că credințele noastre inconștiente ar putea fi servitoarele credințelor conștiente, prin aceea că primele ne abat de la dovezile care le-ar putea submina pe cele din urmă, dar aceasta nu este nimic altceva decât o fabulație fără susținere. Tot într-o manieră speculativă, ne putem imagina că „inconștientul" este capabil să determine strategii indirecte (un pas înapoi, doi pași înainte), de exemplu, făcând ca un copil să se rănească doar pentru a atrage atenția părinților. Aceste sugestii fac inconștientul prea asemănător cu gândirea conștientă prin aceea că este considerat capabil de a avea o reprezentare a viitorului (capitolul 6) și a acțiunilor și intențiilor altor oameni (capitolul 19). Stările mentale de care nu suntem conștienți pot cauza acțiuni spontane, cum ar fi aceea de a răspunde „nu, nu" în locul unui simplu „nu", dar nu cunosc dovezi care să susțină ideea că ar putea cauza comportamente care țin de rațiunea instrumentală.

◈◈◈

Notă bibliografică

Dovezi pentru fixarea evaluărilor de probabilitate sunt date în A. Tversky și D. Kahneman, „Judgment under uncertainty: Heuristics and biases", *Science*, 185 (1974), pp. 1124-1131. Dovezi pentru fixarea preferințelor sunt date în D. Ariely, G. Loevenstein și D. Prelec, „Coherent arbitrariness", *Quarterly Journal of Economics*, 118 (2003), pp. 73-105. Referirea la Episcopul Andrewes se găsește în A. Nicolson, *God's Secretaries* (New York, HarperCollins, 2003), iar cea la taumaturgie în M. Bloch, *Les rois thaumaturges* (Paris, Armand Colin, 1961). O bună discuție despre sentimentalitate se găsește în M. Tanner, „Sentimentality", *Proceedings of the Aristotelian Society*, n.s. 77 (1976-1977), pp. 127-147. Un experiment extins (pe Internet) privitor la prejudecățile inconștiente este redat în B. Nosek, M. Banaji și A. Greenwald, „Harvesting implicit group attitudes and beliefs from a demonstration website", *Group Dynamics*, 6 (2002), pp. 101-115. Pentru un argument în favoarea ideii că Freud a făcut inconștientul să pară prea asemănător cu gândirea conștientă, vezi L. Naccache, *Le nouvel inconscient* (Paris, Odile Jacob, 2006).

BCU IAȘI - CENTRU UNIVERSITAR

Capitolul 4

MOTIVAȚIILE

Capitolul de față și cele două care urmează vor fi dedicate tipurilor de motivație. În acest capitol mă voi rezuma la observații cu caracter general. În capitolul următor mă voi concentra pe două chestiuni particulare, egoism *versus* altruism și miopie temporală *versus* hipermetropie temporală. Acestea sunt două probleme complementare, ultima fiind un fel de versiune intertemporală a celei dintâi, contrastul interpersonal. După cum vom vedea, cele două probleme sunt și puternic înrudite, în sensul că hipermetropia temporală poate *mima* altruismul.

Setul motivațiilor umane este o „plăcintă" care poate fi împărțită în moduri foarte diferite. Cu toate că niciunui mod nu i se poate conferi statut de canon, există patru abordări pe care eu le consider utile. Prima propunere implică un continuum de motivații, a doua și a treia oferă o trihotomie, iar a patra o simplă dihotomie. Clasificările sunt întrucâtva asemănătoare, dar și diferite într-un sens foarte interesant, ceea ce ne permite să facem lumină asupra unui anumit comportament din mai multe perspective.

De la visceral la rațional

Pe 11 septembrie 2001, unii dintre cei care se găseau în clădirea World Trade Center au sărit pe geam din cauza căldurii copleșitoare. „Nu ar trebui să vedem asta ca pe o alegere în adevăratul sens al cuvântului", a spus Louis Garcia, șeful departamentului de pompieri al New Yorkului. „Dacă pui oameni lângă o fereastră și îi supui unei asemenea temperaturi, sunt șanse mari ca mulți dintre ei să se simtă constrânși să sară." Pur și simplu nu le-a fost dată o alternativă de altă natură. Subiectiv, aceasta poate fi experiența acelora care beau apă sărată din mare atunci când nu au la dispoziție apă dulce. Probabil sunt conștienți de faptul că a bea chiar și foarte puțină apă sărată înseamnă a o lua pe un drum foarte periculos: cu cât bei mai mult, cu atât ți se face mai sete. Dar tentația poate fi, pentru unii, irezistibilă. La fel poate fi și experiența de a râvni substanțe care creează dependență. Un scriitor din secolul al XVIII-lea,

BUCHAREST UNIVERSITY LIBRARY

Benjamin Rush, ne furnizează o descriere dramatică: „Atunci când [un băutor înrăit] a fost rugat stăruitor de unul dintre prieteni să se lase de băut, i-a spus acestuia: «Dacă s-ar afla un butoi de rom într-un colț al camerei și dacă un tun ar trage permanent ghiulele între mine și butoi, asta tot nu m-ar opri să trec prin fața tunului pentru a ajunge la rom.»" Dorința sexuală poate fi, de asemenea, atât de copleșitoare încât să reducă la tăcere orice preocupare pentru prudență.

Unele emoții pot fi atât de puternice încât să înăbușe orice alte considerații. Sentimentul rușinii, de exemplu, poate fi insuportabil de dureros, după cum stă mărturie sinuciderea din 1996, a unui amiral american de marină care urma să fie demascat ca nemeritând medaliile pe care le primise sau cele șase sinucideri din 1997 în rândul francezilor care fuseseră demascați drept consumatori de materiale pornografice cu minori. Mânia poate fi și ea covârșitoare, după cum s-a văzut în cazul lui Zinedine Zidane pe 9 iulie 2006, în ultimele minute ale finalei Campionatului Mondial de Fotbal, când fotbalistul a lovit un adversar italian cu capul în piept ca răzbunare pentru o provocare, făcând acest gest în fața a șaptezeci de mii de oameni aflați pe stadion și a unei audiențe estimate, la nivel mondial, de un miliard de telespectatori. Dacă s-ar fi oprit o fracțiune de secundă pentru a se gândi la ceea ce face, poate și-ar fi dat seama că un asemenea gest nu poate aduce decât înfrângerea echipei și ruinarea propriei sale reputații.

Poate cu excepția impulsului de a sări din clădirea World Trade Center, ne îndoim că aceste dorințe ar fi literalmente irezistibile, așa cum un bolovan care a luat-o la vale nu mai poate fi oprit din cădere. (Un impuls de a adormi poate fi irezistibil, dar faptul de a adormi nu este o acțiune; și, de aceea, încercările de a te opune nu duc la rezultatul dorit.) Dependenții sunt întrucâtva sensibili la costuri: consumă mai puțin atunci când prețurile cresc.[1] Cei aflați într-o barcă de salvare se pot opri unii pe alții să bea apă sărată. Cu siguranță, poți opune rezistență impulsului de a te sinucide din rușine sau tentației sexuale. Din cauza intensității lor, aceste pofte viscerale se află, cu toate acestea, la o extremă a spectrului motivațiilor umane. Ele au potențialul, nu întotdeauna realizat, de a bloca deliberarea, compensațiile și chiar alegerea.

La cealaltă extremă găsim o paradigmă – sau, mai degrabă, o caricatură – a agentului rațional care nu este perturbat de factori viscerali, inclusiv de emoție. Acesta acționează doar după ce a cumpănit atent – dar nu mai atent decât este necesar în respectivele circumstanțe – toate consecințele opțiunilor disponibile. Un general rațional, un director general de companie sau un doctor sunt mai preocupați doar de găsirea celor mai bune mijloace pentru a realiza un scop obiectiv, cum ar fi câștigarea unui război, maximizarea profitului sau salvarea unei vieți. Dorințele subiective și

[1] Acest lucru se poate întâmpla și pentru că bugetul nu le permite să consume la același nivel (capitolul 9).

reacțiile viscerale nu au ce căuta în această ecuație. Cu toate că există (după cum vom vedea în capitolul 11) un concept de raționalitate mult mai larg decât această idee fără viață, acest lucru poate servi ca punct de referință.

Un exemplu care să ilustreze distincția dintre motivația viscerală și cea rațională este furnizat de diferența dintre *frica* viscerală și frica din precauție. Deși, în mod obișnuit, considerăm frica o emoție, ea poate fi doar un complex dorință-credință. Când spun „Mă tem că s-ar putea să plouă", vreau să spun *doar* că cred că va ploua și că mi-aș dori să nu plouă. Dacă „frica" inspiră acțiune, ca atunci când îmi iau umbrela pentru a mă proteja de ploaie, aceasta este o paradigmă a comportamentului rațional (capitolul 11). Niciuna dintre proprietățile tipice ale emoțiilor (capitolul 7) nu apare. Frica viscerală, prin contrast, poate determina o acțiune care nu este rațională din punct de vedere instrumental. De exemplu, s-a calculat că 350 de americani și-au pierdut viața în accidente de mașină pentru că au evitat să folosească avionul după incidentele din 11 septembrie 2001. Dacă ar fi luat avionul, ar fi fost încă în viață. Se pare, însă, că același calcul nu este valabil și pentru Spania, unde rata mortalității nu a crescut din cauza faptului că, după incidentele din Madrid de pe 11 martie 2004, oamenii au trecut de la mersul cu trenul la mersul cu mașina. E posibil ca, din cauza seriei îndelungi de atentate venite din partea organizației separatiste basce ETA, populația să fi dezvoltat o atitudine de frică de prevedere în locul unei frici viscerale în raport cu bombardamentele teroriștilor. Pentru ei, atentatele teroriste se poate să fi fost doar un risc printre altele, asemănător – chiar dacă mult mai periculos – cu riscul de a ploua.

Între extremele continuumului visceral-rațional regăsim un comportament parțial motivat de factori viscerali, dar care este sensibil întrucâtva și la considerații care privesc costurile și beneficiile. Cineva poate căuta să se răzbune (o dorință viscerală), însă poate să aștepte până la momentul potrivit, când își poate prinde adversarul nepregătit (o preocupare precaută). Dacă își provoacă inamicul la un duel (așa cum cer normele onoarei), poate lua în secret lecții de scrimă (o practică dezonorantă, dar utilă). Dacă unei persoane i se face o ofertă care este, în același timp, nedreaptă, dar avantajoasă, în sensul că acceptarea ofertei este mai benefică decât respingerea ei, persoana ar putea să o accepte sau nu în funcție de puterea interesului în raport cu puterea indignării.[1] În cazuri mai complexe, un factor visceral îl poate contrabalansa pe altul. Dorința pentru o aventură extraconjugală poate fi neutralizată de sentimentele de vină. Un imbold de a fugi generat de frică poate fi anulat sau înlocuit de un imbold de a lupta cauzat de mânie.

---

[1] Într-un caz studiat pe larg (vezi capitolul 20), respingerea implică faptul că persoana căreia i s-a făcut oferta suferă și de un factor care va spori motivul indignării.

Interes, rațiune și pasiune

În analiza motivațiilor umane, moraliștii francezi din secolul al XVII-lea obișnuiau să facă o distincție fructuoasă între interes, rațiune și pasiune. Interesul presupune urmărirea avantajului personal: fie bani, fie faimă, fie putere, fie mântuire. Chiar și acțiunile prin care ne ajutăm copiii intră în categoria urmării unui interes întrucât viața noastră este puternic legată de a lor. Un părinte care își trimite copiii la o școală privată scumpă, unde aceștia se pot bucura de cea mai bună educație, nu își sacrifică interesul, ci îl realizează. Putem considera că pasiunile includ emoțiile, cât și alte impulsuri viscerale, cum ar fi foamea, setea sau poftele sexuale ori acelea cauzate de dependență. Anticii includeau și stările de nebunie în aceeași categorie generală întrucât, la fel ca emoțiile, acestea sunt involuntare, spontane și subminează deliberarea rațională. În general, printre pasiuni putem include și stările de intoxicare. Din punctul de vedere al legii, mânia, beția și nebunia au fost de cele mai mult ori tratate în același mod.

Rațiunea este o idee ceva mai complicată. Moraliștii o foloseau cel mai des (la fel cum o voi face și eu aici) în relație cu dorința de a încuraja binele public mai degrabă decât scopurile private. Uneori, foloseau termenul pentru a se referi la motivații pe termen lung (de prevedere), ca fiind opuse acelora pe termen scurt (mioape). Ambele idei pot fi rezumate sub titlul de *imparțialitate*. În conceperea politicilor publice, indivizii ar trebui tratați imparțial, astfel încât nu ar trebui să existe indivizi sau grupuri care să se bucure de un tratament privilegiat. Și indivizii pot acționa cu o astfel de motivație. Părinții își pot sacrifica interesul trimițându-și copiii la o școală publică întrucât cred în egalitatea de șanse. În același timp, decidenții politici, precum și indivizii privați ar trebui să trateze rezultatele care apar la momente diferite în timp într-un mod imparțial, conferindu-le aceeași pondere în deciziile adoptate curent, și nu privilegiind rezultatele din viitorul apropiat. De fapt, unii moraliști susțineau că preocuparea pentru interesul pe termen lung are tendința de a încuraja binele public. La Convenția Federală din Philadelphia, de exemplu, George Mason a susținut că

> ar trebui să ne preocupe drepturile fiecărei clase de oameni. Adeseori m-a nedumerit indiferența afișată de clasele superioare ale societății față de acest dictat al umanității și al politicii; ținând cont că, oricât de mare ar fi bunăstarea acestora sau cât de înaltă condiția lor în decursul a doar câțiva ani, posteritatea lor va fi cu siguranță distribuită spre clasele de jos ale Societății. Orice motiv egoist, prin urmare, orice atașament față de familie ar trebui să recomande un sistem de adoptare a politicilor astfel încât acesta să nu presupună dispoziții mai atente pentru drepturile și fericirea claselor mai înalte decât pentru acelea ale claselor mai joase de Cetățeni.

Orice formă de imparțialitate are mai multe grade. Puterea preocupării pentru ceilalți tinde să varieze invers proporțional nu doar cu distanța genealogică, dar și cu

cea geografică. Similar, chiar și indivizii prudenți pun o mai mare greutate pe viitorul apropiat decât pe acela foarte îndepărtat, un fapt care este explicat doar parțial prin posibilitatea de a muri înainte să apuce să se bucure de acel viitor îndepărtat.

Ca un exemplu al modului în care comportamentul poate fi înțeles în termenii oricăreia dintre aceste motivații, putem cita o scrisoare din 1783 trimisă de la New York de Robert Livingston lui Alexander Hamilton; aceasta conține observații cu privire la persecutarea acelora care au luat partea britanicilor în Războiul pentru Independență:

> Regret, la rândul meu, alături de tine, caracterul violent al persecuțiilor care domină aici și mă tem de consecințele asupra bogăției, comerțului și liniștii viitoare a statului. Eu sunt cu atât mai îndurerat de ce se întâmplă întrucât se pare că toate violențele nu au nimic de-a face cu motive patriotice mai nobile. Pentru câțiva, e vorba doar de un spirit orb al *răzbunării și resentimentului*, dar, pentru cei mai mulți, este vorba doar de cel mai *sordid interes*.

Expresiile pe care le-am scos în evidență cu caractere italice corespund rațiunii, emoțiilor și, respectiv, interesului. Adjectivele sunt foarte ilustrative: rațiunea este pură, pasiunea este oarbă, iar interesul este sordid. Voi reveni asupra semnificațiilor acestor evaluări.

Sine, eu, supraeu

În analiza motivațiilor umane, Freud a sugerat și trei forme de bază, fiecare dintre ele legată de un subsistem separat al minții. Cele trei sisteme sunt: sinele, eul și supraeul, care corespund principiului plăcerii, principiului realității și, respectiv, conștiinței. Sinele și supraeul reprezintă pulsiunile, respectiv controlul pulsiunilor, în timp ce eul, „neajutorat în ambele direcții, (...) se apără în van, atât împotriva instigărilor sinelui ucigaș, cât și împotriva reproșurilor conștiinței punitive". Într-un enunț edificator din același eseu („Eul și Sinele"), Freud scrie că eul este „o creatură nefericită, slujind trei stăpâni și amenințată, pe cale de consecință, de trei pericole, venite dinspre lumea exterioară, dinspre libidoul sinelui și dinspre severitatea supraeului". Și totuși, nici această formulare nu surprinde pe de-a-ntregul ceea ce cred eu că reprezintă nucleul util al ideii lui Freud, anume afirmația potrivit căreia eul, în timp ce navighează prin lumea exterioară (principiul realității), trebuie să ducă un război pe două fronturi: împotriva pulsiunilor sinelui (principiul plăcerii) și împotriva controlului punitiv sever al pulsiunilor exercitat de supraeu (conștiință).[1]

[1] Putem combina două metafore ale lui Freud: eul este ca un călăreț pe un cal nărăvaș (sinele), călărit la rândul său de un incubus (supraeul).

Această propunere este originală, profundă și adevărată. Îi lipsește doar un mecanism: de ce nu ar putea chiar eul să exercite orice formă de control al pulsiunilor de care este nevoie? De ce moralitatea și conștiința iau de cele mai multe ori forma unor reguli rigide? Trebuie să stipulăm existența unor funcții mentale distincte și cvasi-autonome? Abia munca de pionierat desfășurată de George Ainslie a dat un răspuns adecvat la aceste întrebări. Voi discuta viziunea acestuia în capitolul 13. Aici vreau doar să atrag atenția asupra faptului că multe pulsiuni trebuie ținute la distanță din cauza daunelor *cumulative* pe care le pot provoca dacă nu sunt controlate.[1] Să luăm un exemplu oarecare, băutul sau mâncatul în exces, dezmățul sau amânarea (ca, de exemplu, să nu-ți faci temele la timp); acestea nu sunt neapărat foarte dăunătoare agentului. Daunele apar dacă excesele (sau eșecurile) sunt repetate. Prin urmare, punctul în care trebuie să se concentreze controlul pulsiunilor nu trebuie să fie ocazia singulară, din moment ce persoana își poate spune că va începe o viață nouă mai bună chiar de mâine. Controlul pulsiunilor trebuie să se adreseze faptului că pulsiunea se va ivi într-un mod predictibil într-un număr indefinit de ocazii. Soluția poate fi găsită prin modificarea cadrului în care vedem problema, astfel încât eșecul de a controla o pulsiune într-o anumită ocazie este văzut ca un factor de predicție a eșecului de a impune controlul în ocazii ulterioare. „Da, pot amâna controlul pulsiunilor până mâine fără să atrag prejudicii importante sau riscuri, dar de ce ar fi ziua de mâine diferită cu ceva de ziua de azi? Dacă eșuez acum, voi eșua și mâine." Prin stabilirea unui *efect de domino intern*, care mărește miza, agentul poate căpăta o motivație să își controleze pulsiunile, care ar lipsi dacă ar lua în considerare fiecare zi în parte. Reversul monedei este că trebuie exercitat un control continuu, după cum spun moraliștii victorieni: „Să nu admiți nici măcar o singură excepție."

Ținând cont de consecințe

În cele din urmă, motivațiile pot fi consecinționiste sau non-consecinționiste, adică pot fi orientate fie spre rezultatul unei acțiuni, fie spre acțiunea în sine. O mare parte a comportamentului economic este de natură pur consecinționistă. Atunci când oamenii economisesc pentru bătrânețe sau când brokerii de la bursă cumpără și vând acțiuni, nu atașează nicio valoare intrinsecă – fie pozitivă, fie negativă – acestor acțiuni în sine; le pasă doar de rezultate. Dimpotrivă, pacifistul înverșunat, care refuză să meargă la război chiar și împotriva celui mai malefic dușman, nu ține cont de consecințele comportamentului său. Ce contează pentru el este că unele acțiuni sunt

[1] Există, de asemenea, și un *risc* cumulativ. Șansele unor consecințe nedorite în urma sexului neprotejat poate că sunt mici de fiecare dată, dar riscul pe termen lung poate fi considerabil. Într-o excursie, șansele de a fi rănit într-un accident de mașină când nu porți centura de siguranță sunt foarte mici, dar probabilitatea raportată la o viață întreagă este de aproximativ 1 din 3.

interzise necondiționat, ca, de exemplu, a lua viața unui om. Nu e ca și cum acesta *nu ar fi conștient* de consecințe, cum se poate întâmpla cu acțiunile pornite din emoție, ci, pur și simplu, consecințele sunt irelevante pentru ceea ce face.

Politicile publice pot fi adoptate, de asemenea, recurgând la oricare dintre cele două tipuri de motivație. Un politician poate adopta principiul „Cine găsește un obiect pierdut poate să-l păstreze" (de exemplu, în legislația referitoare la patente), presupunându-se că, dacă se atribuie un drept de proprietate persoanei care descoperă o nouă resursă de valoare, acest lucru poate încuraja descoperirea mai multor resurse de valoare. Argumentul acesta este consecinționist. Un argument non-consecinționist pentru aceeași politică ar fi că persoana care descoperă o nouă resursă, fie o bucată de pământ, fie un tratament împotriva cancerului, are un *drept* natural de proprietate asupra descoperirii. Pentru alt exemplu, putem lua în considerare Discursul XXXI al lui Dion Hrisostomos împotriva obiceiului locuitorilor din Rodos de a refolosi statuile vechi din bronz pentru a-i cinsti pe binefăcătorii orașului: acest obicei, argumenta el, înseamnă atât a viola drepturile acelora în onoarea cărora fuseseră ridicate statuile inițiale, cât și a descuraja potențiali noi binefăcători, care știau că statuile ridicate în cinstea lor vor fi cât de curând refolosite în favoarea altcuiva. Argumentele consecinționiste pot (părea) să justifice niște măsuri dure împotriva teroriștilor, chiar dacă pașii care trebuie urmați pentru a le pune în practică violează valorile non-consecinționiste asociate cu drepturile omului și libertățile cetățeanului.[1]

Un caz special de motivație non-consecinționistă este principiul care ar putea fi numit kantianism banal, imperativul categoric sau gândirea magică: *fă ceea ce ar fi mai bine dacă toată lumea ar face la fel.* Într-un sens, principiul acesta este legat de consecințe, din moment ce agentul face acel ceva care ar duce la cel mai bun rezultat, în cazul în care toată lumea ar face același lucru. Acestea nu sunt, însă, consecințele acțiunilor *sale*, ci ale unui set ipotetic de acțiuni, ale sale și ale altora. Într-un caz dat, a acționa în baza acestui principiu poate avea consecințe dezastruoase pentru toți dacă ceilalți *nu* se comportă la fel. În arena internațională, dezarmarea unilaterală este un astfel de exemplu.

Un alt caz este următorul principiu al eticii iudaice. Să presupunem că dușmanul se află la ușă și spune: „Dați-mi-l pe unul dintre ai voștri pentru a fi ucis și ceilalți vor fi cruțați; dacă refuzați, veți muri cu toții." Talmudul cere ca, într-un astfel de caz, toți evreii să se lase uciși mai degrabă decât să fie numit unul dintre ei pentru a fi ucis astfel încât ceilalți să-și păstreze viața. Dar dacă dușmanul spune „Dați-mi-l pe Petru" în aceleași condiții, este acceptabil ca acesta să fie predat. Nu există o interdicție privitoare

[1] Acest „părea" pus între paranteze reflectă o posibilă funcționare a „psihologiei tiraniei" (capitolul 2). O dilemă clasică a intimidării este că *ura* pe care o inspiră poate, într-un final, să contrabalanseze *frica* pe care trebuia să o provoace.

la moartea unei persoane pentru a salva viața altora, ci doar la faptul de a alege cine va fi acea persoană. Romanul *Alegerea Sofiei* ne pune în fața unei dileme asemănătoare.

Normele sociale (capitolul 22) ne pun la dispoziție un caz special al comportamentului non-consecinționist, cu o particularitate importantă. Normele sociale le spun oamenilor ce să facă, de exemplu, să se răzbune pentru o insultă sau să nu mănânce un copil fiert în laptele mamei sale, nu pentru că ar exista anumite rezultate dezirabile care ar trebui înfăptuite, ci pentru că acțiunea este obligatorie în sine.[1] Cu toate că nu au scopul de a *produce* un anumit rezultat, se poate considera că asemenea acțiuni sunt întreprinse pentru *a împiedica* un anumit rezultat, și anume mustrarea din partea celorlalți pentru a nu le fi înfăptuit. Și totuși, ne putem întreba dacă mustrările sunt și ele întreprinse din motive consecinționiste asemănătoare. În general, voi argumenta că nu. Mai mult, atunci când oamenii sunt răniți de acțiunile celorlalți, se răzbună chiar și în cadrul interacțiunilor singulare desfășurate în deplin anonimat, ca, de exemplu, cea pe care o pot furniza circumstanțele unui experiment. Pentru că interacțiunea este singulară, aceștia nu au nimic de câștigat din întâlniri ulterioare și, pentru că se desfășoară în anonimat, aceștia nu au a se teme de mustrările unor terți. Mă voi întoarce asupra acestor experimente în capitolele următoare.

Chiar și pentru un non-consecinționist declarat, consecințele pot conta dacă sunt suficient de importante. Să luăm în considerare un principiu pe care cei mai mulți l-ar vedea drept necondiționat: interdicția de a tortura copii mici. Într-un scenariu care implică o „bombă cu ceas", ne putem imagina că o condiție necesară și suficientă pentru a preveni detonarea unui dispozitiv nuclear în centrul Manhattanului este aceea de a-l tortura pe copilul mic al unui terorist în fața acestuia. *Dacă* am putea face scenariul credibil, mulți non-consecinționiști ar putea să fie de acord cu tortura. Alții ar putea spune că, din moment ce niciodată nu vom avea astfel de circumstanțe în viața reală, interdicția absolută rămâne în picioare. Iar alții ar interzice tortura chiar dacă scenariul ar avea loc. Sarcina mea aici nu este de a da argumente în favoarea unei anumite concluzii, ci să fac o observație de natură empirică: situațiile din viața reală au rareori o miză atât de mare încât să îl facă pe non-consecinționist să ia în considerare consecințele comportamentului său. Dacă miza este suficient de mare, s-ar putea ca acesta să-și abandoneze principiile, dar, întrucât situația nu apare, nu ne putem da seama cu

[1] Cu privire la regulile hranei cușer, pentru care interdicția de a mânca un copil fiert în laptele propriei mame este doar o ilustrație istorică, s-a crezut pentru o vreme că își găsesc justificarea pe motive de igienă. Astăzi, din câte îmi dau seama, sunt recomandate pe motiv că este bine pentru cineva să facă un lucru în același timp dificil și lipsit de sens. Această idee pare a încorpora eroarea produselor secundare, pe care o voi discuta mai târziu: un comportament care este justificat *doar* prin efectele sale de a întări caracterul nu va avea nici măcar aceste efecte. Nu știu, însă, câți dintre aceia care urmează această regulă o fac din acest motiv.

certitudine dacă avem de a face cu un troc greu de obținut sau cu un refuz total de a face troc.

Aceste patru abordări ale motivației surprind fenomene identice. Factorii viscerali, pasiunile și principiul plăcerii au, în mod evident, destul de multe în comun. Cel din urmă este aplicabil unui număr mai mare de cazuri întrucât implică evitarea durerii, precum și căutarea plăcerii. Când studenții amână să se apuce de învățat, nu fac acest lucru neapărat pentru că ard de nerăbdare să facă altceva. De cele mai multe ori, aceștia urmează calea minimei rezistențe. Supraeul, rațiunea și motivațiile non-consecinționiste au, de asemenea, anumite trăsături în comun. Cu toate că nu toate sistemele morale sunt rigide și implacabile, unele sunt. Teoria morală a lui Kant este un exemplu faimos. (De fapt, filosofia sa morală se poate să fi izvorât din regulile private pe care le-a instituit pentru a-și controla propriile pulsiuni, ca, de exemplu, maxima de a nu fuma niciodată mai mult de o pipă după micul dejun.) În același timp, moralitatea se poate ridica deasupra rigidității în cazul indivizilor care nu au nicio aversiune față de ambiguitate. De fapt, toleranța față de ambiguitate este considerată, de cele mai multe ori, punctul de referință al unui eu sănătos. Spre deosebire de aceasta, relația dintre raționalitate, interes, eu și consecinționism este mult mai subtilă. Ar fi absurd să susținem că punctul de referință al unui eu sănătos este urmărirea rațională a propriului interes.

A vrea și a dori

De cele mai multe ori, gândim motivațiile sub forma faptului de *a vrea* ca o anumită stare de fapt *să se întâmple*. Acestea pot lua, de asemenea, și forma de *a dori* ca o anumită stare de fapt *să se întâmple*. Distincția dintre a vrea și a dori este importantă dacă ne referim la componenta motivațională a emoției (capitolul 7). Emoțiile pot fi însoțite fie de a vrea să faci ceva, fie de a dori ca un anumit lucru să se întâmple. În cazul *mâniei*, impulsul lui A de a se răzbuna pe B nu poate fi satisfăcut de faptul că altcineva, C, îi face lui B ceea ce plănuise A să-i facă sau de faptul că B suferă un accident. Aici nu contează doar rezultatul în sine, acela că B suferă, ci contează ca B să sufere în urma unei acțiuni a lui A. De asemenea, în cazul sadismului contează să îl faci pe celălalt să sufere, nu doar ca acesta să sufere pur și simplu. Dimpotrivă, în cazul *urii* contează ca persoana sau grupul care constituie obiectul urii să dispară de fața pământului, indiferent dacă acest lucru se întâmplă ca urmare a acțiunii aceluia care urăște sau a altcuiva. Și în cazul *ranchiunei* contează suferința celuilalt, nu neapărat provocată de persoana ranchiunoasă. De fapt, o persoană ranchiunoasă poate chiar să se ferească să provoace în mod activ suferința celuilalt, nu din teama de a nu fi văzută, ci pentru că acest lucru ar fi incompatibil cu imaginea de sine. Acest lucru este chiar mai clar în cazul *invidiei*. Multora le-ar face plăcere să-și vadă rivalii aduși în sapă de lemn și nu ar face nimic să împiedice acest lucru, dacă ar putea, însă nu ar face nimic

concret în dauna acelor rivali, chiar dacă planul nu atrage după sine niciun fel de costuri sau riscuri.[1] O persoană care nu ar da foc niciodată casei vecinului ar putea foarte bine să nu cheme pompierii dacă ar vedea casa în flăcări.

Gândirea dezideravă (capitolul 7) nu se bazează pe a vrea, ci pe a dori. În unele cazuri, agentul se eschivează de la dificila sarcină de a face lumea să se conformeze dorințelor sale și poate adopta, în schimb, calea ușoară a asumării unei credințe corespunzătoare despre lume. Dacă doresc să fiu promovat, dar sunt reticent când vine vorba de a face un efort în acest sens, mă pot baza pe semne de mică importanță prin care să mă conving pe mine însumi de faptul că promovarea este iminentă. În alte cazuri, acțiunea asupra lumii nu se prezintă ca o opțiune viabilă. Nu pot cauza reciprocitatea sentimentului de iubire sau însănătoșirea copilului meu. În astfel de cazuri, fie mă pot angaja în iluzii încântătoare, fie pot accepta realitatea. O altă distincție este aceea între cazurile în care iluziile nu au consecințe ulterioare pentru acțiune și acelea în care iluziile sunt folosite ca premise pentru acțiune. Mă pot amăgi cu faptul că o femeie pe care o cunosc nutrește în secret o pasiune puternică pentru mine și, totuși, nu sunt în stare să-i fac vreo propunere, fie pentru că sunt constrâns de moralitate (sau de interesul personal), fie pentru că credința cu care mă amăgesc este întreținută în bună măsură pentru valoarea ei de consum.[2] În unele cazuri, amăgirea poate fi exprimată deschis obiectului dorinței, cum s-a întâmplat cu o secretară a lui John Maynard Keynes, care i-a spus că nu a putut să nu remarce marea pasiune pe care acesta o nutrește pentru ea. Și-a distrus viața în felul acesta.

Stări care, în esență, sunt produse derivate

Un factor care complică distincția dintre a vrea și a dori este acela că, în anumite cazuri, pot obține X făcând A, dar numai dacă fac A pentru a obține Y. Dacă muncesc din greu pentru a explica baza neurofiziologică a emoției și am succes, pot câștiga o bună reputație. Dacă mă implic într-o muncă intensă pentru o anumită cauză politică, pot descoperi la finalul întregului demers că m-am ales cu o anumită „reputație". Dacă știu să cânt bine la pian, îi pot impresiona pe ceilalți. Toate aceste beneficii indirecte parazitează scopul principal al activității. Dacă motivația mea ca om de știință este aceea de a câștiga o anumită reputație, e puțin probabil că o voi dobândi. A intra într-o mișcare politică *doar* de dragul unor efecte pozitive asupra conștiinței sau pentru a căpăta o anumită experiență înseamnă eșec sigur sau succes doar prin accident.

---

[1] Sigur, unii oameni invidioși nu au astfel de probleme. Aceștia pot trăi într-o societate în care invidia nu este percepută drept ceva foarte rușinos sau poate ei înșiși sunt nerușinați.

[2] Putem remarca pentru un punct de referință ulterior (capitolul 23) că gândirea antidezideravă nu poate avea valoare de consum. Prin urmare, atunci când se manifestă, cel mai probabil va servi ca premisă a acțiunii.

După cum a observat Proust, un muzician „poate să-și trădeze uneori adevărata vocație pentru glorie, dar, atunci când caută gloria în acest mod, se îndepărtează tot mai mult de ea și o descoperă doar când îi întoarce spatele". Conștiința de sine are efecte nedorite asupra acțiunii. Tot Proust a spus că, deși „modul cel mai eficient de a te face dorit este să fii greu de găsit", nu ar sfătui pe nimeni să procedeze așa întrucât „o astfel de metodă de a obține succesul în societate funcționează doar dacă nu este adoptată în acest scop".

Gloria în muzică sau succesul social pot fi incluse în categoria *stărilor care sunt în esență produse derivate*, stări care nu pot fi atinse prin acțiuni motivate doar de dorința de a le atinge. Acestea sunt stări care pot *apărea*, dar care nu pot fi *făcute să survină* în mod intenționat sau printr-o simplă decizie. Printre aceste stări se mai numără: dorința de a uita, dorința de a crede sau dorința de a dori (ca, de exemplu, dorința de a depăși impotența sexuală), dorința de a dormi, dorința de a râde (nu te poți gâdila pe tine însuți) sau dorința de a nu mai fi bâlbâit. Încercările de a realiza astfel de dorințe sunt, cel mai probabil, fără succes și chiar pot înrăutăți situația. Este un loc comun printre moraliști și romancieri că hedonismul intenționat nu poate deveni o realitate[1] și că nimic nu face o amintire mai prezentă decât încercarea de a o uita. Cu toate că ne putem dori să atingem astfel de stări, trebuie să ne ferim să vrem să le atingem.

Pentru foarte mulți oameni este importantă *mântuirea* (în viața de apoi) și *izbăvirea* (de păcatele comise). Aceștia pot crede că ating aceste scopuri prin acțiune. A muri ca martir în lupta împotriva necredincioșilor poate fi un pașaport spre Rai, conform credinței unora. A lupta împotriva naziștilor după ce ai colaborat inițial cu ei poate izbăvi fărădelegea. Și totuși, dacă aceste acțiuni sunt întreprinse cu *scopul* de a atinge mântuirea sau izbăvirea, ele ar putea eșua. În teologia catolică, intenția de a cumpăra un loc în Rai prin martiriu voluntar ar fi o instanță a păcatului simoniei. Unii experți în studii islamice aduc o critică similară atacurilor teroriste sinucigașe, motivate de credința că autorii lor vor obține un loc privilegiat în Paradis. Montaigne scrie că, atunci când spartanii „trebuiau să decidă care dintre ei va primi onorurile pentru cele mai mari fapte de vitejie din ziua respectivă, aceștia au decis că Aristodemus a înfruntat primejdiile în modul cel mai curajos, însă nu i-au conferit onorurile întrucât dârzenia sa își avea imboldul în dorința de a se curăța de reproșurile pe care și le atrăsese după bătălia de la Termopile". Magnatul francez de presă Jean Prouvost, care a colaborat cu forțele naziste în timpul ocupației din Franța, a încercat să își îndrepte greșeala

[1] În ultimul volum din *În căutarea timpului pierdut*, Proust, reflectând probabil asupra propriei sale vieți, scrie că a căuta fericirea, o întreprindere zadarnică, poate duce la o anumită înțelegere a condiției umane și poate oferi „o anumită bucurie". Căutarea unor stări care sunt esențialmente produse derivate le poate provoca indirect pe acestea; dacă un copil pune pe cineva să râdă, persoana poate chiar să izbucnească în râs ca reacție la o cerință ridicolă.

donând o sumă mare de bani Rezistenței atunci când era deja clar că germanii vor pierde războiul. După eliberare, Curtea Supremă i-a oferit un *non-lieu* (o decizie judecătorească prin care se suspendă, se anulează sau se retrage un caz fără ca acesta să mai fie adus în instanță), un lucru pe care probabil că spartanii nu l-ar fi făcut.[1]

Presiune versus *atracție*

De ce părăsesc oamenii o anumită țară în favoarea alteia? De ce părăsesc cadrele universitare o universitate în favoarea alteia? De multe ori, răspunsurile la aceste întrebări sunt clasificate la categoria „presiune *versus* atracție". Cineva poate emigra fie pentru că situația de acasă este insuportabilă, fie pentru că situația din străinătate este irezistibil de atrăgătoare – cel puțin acesta este modul obișnuit de a privi problema. Cu toate acestea, în multe situații suntem induși în eroare dacă gândim astfel. În mod obișnuit, oamenii emigrează deoarece *compară* situația de acasă cu cea din străinătate și pot ajunge la concluzia că diferența este suficient de mare încât să justifice plecarea, chiar ținând cont de costurile pe care aceasta le presupune.[2] Și totuși, distincția dintre motivele de presiune și cele de atracție poate avea sens dacă primele sunt mai apropiate de partea viscerală a continuumului, iar cele din urmă sunt mai aproape de partea rațională. Cei care sunt cuprinși de frică puternică fug uneori din fața pericolului, și nu neapărat înspre siguranță. Singurul lucru pe care îl au în minte este să fugă, nemaiavând răgazul necesar pentru a se gândi dacă nu cumva vor ajunge din lac în puț. În funcție de drog, dependenții pot fi motivați fie de atracția euforiei (în cazul cocainei), fie de presiunea disforiei (în cazul heroinei). Comportamentul suicidal poate, de asemenea, să datoreze mai mult presiunii decât atracției, fiind, mai degrabă, o evadare din fața disperării, și nu o fugă spre ceva anume.

Și modul în care operează normele sociale (capitolul 22) poate fi considerat în termeni de presiune *versus* atracție. Dorința de a excela în moduri aprobate de societate exercită o puternică atracție pentru mulți indivizi, fie că aceștia se luptă pentru *glorie* (a fi cel mai bun) sau pentru *onoare* (a câștiga într-o competiție sau luptă). Alți indivizi sunt mult mai preocupați să evite rușinea atașată violării normelor sociale. În multe societăți există o normă generală care poate fi redată astfel: „Nu-ți pune pielea

---

[1] Probabil că a fost eliberat pentru că Rezistența avea nevoie de bani, iar mai târziu s-a găsit obligată să-și respecte promisiunea tacită a imunității pe care o presupunea acceptarea ceculul.

[2] Această formulare presupune un echilibru între costul emigrării și beneficiile acesteia, ca factori determinanți ai utilității globale a migrării. Cu toate acestea, costul emigrării poate intra în ecuație drept o *constrângere* asupra deciziilor. Dacă cel mai ieftin zbor transatlantic costă mai mult decât suma maximă pe care o poate economisi sau împrumuta un țăran italian sărac, acesta va rămâne în Italia indiferent de cât de bună ar fi situația lui din Statele Unite (capitolul 9).

în saramură." A excela în ceva înseamnă a devia de la regulă, iar devierea este obiectul dezaprobării universale: „Cine se crede?" Puterea relativă a acestor două motivații variază între societăți și pe diverse paliere ale aceleiași societăți. Atena antică este un bun exemplu care ilustrează năzuința competitivă spre excelență.[1] Orașele mici ale societăților moderne ilustrează efectele înăbușitoare ale ostilității față de excelență. Chiar dacă risc să generalizez, presiunea exercitată de rușine pare a fi, în ansamblu, o motivație mai importantă decât atracția exercitată de glorie, dar nu vreau să spun prin aceasta că ultima nu se poate manifesta cu o foarte mare putere.

Conflictul motivațional

Existența unor motivații *care concurează între ele* este bine-cunoscută:

Am nevoie de o carte atât de tare, încât sunt tentat să o fur de la bibliotecă, dar vreau să mă comport moral.

În fața unui golan sunt în același timp temător și furios: vreau să fug, dar și să-l pocnesc.

Vreau ca toți copiii să aibă acces la educație publică, dar vreau ca al meu să meargă la o școală privată, unde va beneficia de cea mai bună educație.

Vreau un candidat care este în favoarea legalizării avortului, dar vreau și unul care să susțină impozite mai mici.

Vreau să fumez, dar și să-mi mențin sănătatea.

Dacă mi se face o propunere avantajoasă, dar incorectă, de genul „ori accepți, ori nu accepți", vreau în același timp să resping propunerea întrucât este incorectă, dar să o și accept întrucât este avantajoasă.

Vreau să fac acte de caritate, dar vreau să-mi promovez și propriul interes.

Sunt tentat de perspectiva unei relații extraconjugale, dar vreau în același timp să-mi păstrez căsnicia.

Cum se poate rezolva conflictul prezent la nivelul acestor motivații? Un răspuns general poate fi următorul: atunci când situația este de tipul „câștigătorul ia tot", prin urmare, nici un compromis (fizic) nu este posibil, câștigă cea mai puternică motivație.[2] Dacă preocuparea pentru propriul meu copil este mai puternică decât grija pentru

---

[1] De pildă, Eschil și-a scris piesele pentru a fi jucate într-un concurs de dramaturgie. Când a fost înfrânt de tânărul Sofocle, amărăciunea lui a fost atât de mare încât a părăsit Atena pentru Sicilia.

[2] Cel puțin în principiu, se poate utiliza un compromis probabilist prin aranjarea unei loterii ponderate între opțiuni, folosind tăria motivelor în favoarea lor. Unele instituții alocă o anumită parte a resurselor, însă una foarte mică, într-un asemenea mod, dar încă nu am întâlnit instanțe de conflicte motivaționale individuale care să fie rezolvate astfel.

educația copiilor în general, atunci îl voi trimite la o școală privată. Dacă preocuparea mea pentru legalizarea avortului este mai mare decât aceea pentru scăderea nivelului de impozitare și niciunul dintre candidați nu favorizează ambele poziții, atunci îl voi vota pe acela care propune legalizarea avortului și creșterea impozitelor. Dacă cineva îmi oferă trei dolari dintr-un fond comun de zece dolari, intenționând să păstreze restul pentru sine, accept. Dacă mi se oferă doar doi dolari, resping oferta.[1] Atunci când este posibil compromisul, motivația mai puternică are un impact mai puternic decât motivația mai slabă. Un fumător poate reduce consumul de țigări de la treizeci de țigări la zece pe zi. Ca o reflectare a tăriei altruismului meu, pot dona 5% din venituri în scopuri caritabile.[2]

Un astfel de răspuns nu este propriu-zis rău, dar este mult prea simplist, din moment ce ideea de „tărie a motivației" este mult mai complicată decât o sugerează aceste exemple date în fugă. O motivație își poate datora tăria simplei sale forțe de natură psihică, acesta fiind sensul în care, de exemplu, motivele viscerale sunt de cele mai multe ori mai puternice decât ceea ce Madison numea „vocea blândă a rațiunii". Și totuși, o motivație puternică poate, de asemenea, să fie puternic aprobată de agent pentru că societatea îi conferă acesteia o valoare foarte mare. Fiecare societate sau cultură este, de fapt, caracterizată de o ierarhie normativă a motivațiilor. Dacă totul merge cum trebuie, o persoană ar face o anumită acțiune pentru motivul A mai degrabă decât pentru motivul B dacă A ocupă un loc mai de frunte în ierarhia socială. Acestea sunt *metamotivațiile*, dorințe care urmează a fi animate de dorințe de un anumit tip.[3] Chiar dacă sunt mai slabe în sens visceral, în final acestea pot ieși învingătoare asupra altor motivații.

Este bine-cunoscut faptul că interesul și pasiunea sunt de foarte multe ori *supuse judecății rațiunii*.[4] După cum a spus Seneca, „rațiunea își dorește ca decizia pe care a luat-o să fie dreaptă; mânia își dorește ca decizia pe care a luat-o să pară cea dreaptă."

---

[1] După cum s-a observat, prin respingerea ofertei mă pot asigura și că ofertantul nu se alege cu nimic.

[2] În capitolul 6 voi discuta un fenomen mult mai paradoxal, „perdantul ia tot", observat în cazul slăbiciunii voinței.

[3] Ideea metamotivațiilor nu are legătură cu conceptul metapreferințelor. Un exemplu pentru cele din urmă ar fi cazul unei persoane care are două ordonări diferite ale preferințelor, una pentru a mânca în loc de a ține regim, alta pentru a ține regim în loc de a mânca, și o metapreferință favorizând-o pe ultima. Urmând intuiția lui La Bruyère că „oamenii sunt foarte vanitoși și detestă, dintre toate lucrurile, să fie considerați astfel", o metamotivație poate fi echivalentă cu preferința de a alege regimul din motive de sănătate decât cu faptul de a avea aceeași preferință din motive de vanitate.

[4] După cum vom vedea în capitolul 12, agenții pot manifesta uneori un respect prea mare față de raționalitate.

Pentru că există numeroase concepții despre rațiune, dreptate și echitate care sună plauzibil, va fi posibil, într-adevăr, de foarte multe ori să prezinți o decizie luată la mânie ca fiind conformă cu rațiunea. Procesele colaboraționiștilor din țările aflate sub ocupație nazistă în al Doilea Război Mondial au fost, în cele mai multe cazuri, înrădăcinate într-o dorință adâncă de răzbunare. Dar, din cauza deferenței față de rațiune, combinată cu dorința de a se delimita de practicile ilegale ale regimului de ocupație, noii lideri și-au prezentat măsurile severe ca fundamentate pe ideea de drept, și nu pe emoție. Cineva poate avea un interes de prim ordin să nu dea bani în scopuri caritabile și o dorință de ordin secund de a nu se vedea pe sine ca motivat doar de interesul personal. Acesta poate adopta filosofia carității din cauza deferenței față de rațiune (capitolul 2), care poate justifica anumite donații mici. Dacă ceilalți dau mult, acesta va adopta o politică utilitaristă, care justifică donațiile mici, iar, dacă ceilalți dau puțin, va adopta o politică bazată pe echitate, care va justifica același comportament.

În asemenea cazuri, rațiunea nu are un rol independent cauzal, ci doar determină o justificare după ce faptul s-a consumat pentru acțiuni decise deja pe alte temeiuri. Conflictul nu este, prin aceasta, rezolvat, ci doar vârât sub preș. În alte cazuri, căutarea unei justificări bazate pe rațiune poate aduce modificări în comportament. Dacă eu adopt o politică a donațiilor bazată pe echitate pentru că ceilalți dau puțin și, subit, aceștia încep să-și mărească donațiile, sunt obligat să mă conformez. Aceeași *nevoie de stimă de sine* care a cauzat justificarea inițială a unui comportament autointeresat prin apel la considerații imparțiale mă împiedică să îmi schimb concepția imparțialității atunci când aceasta nu mai este în favoarea mea. Ne putem gândi că, în *Regele Lear*, atât Ducele Burgundiei, cât și Regele Franței s-au îndrăgostit inițial de Cordelia datorită perspectivei ca aceasta să obțină cea mai mare dotă, însă numai primul a manifestat o grijă cât de mică pentru propria sa imagine de sine, astfel încât să fie în stare să simtă aceeași emoție odată ce sentimentele nu mai coincideau cu interesul. În cazul acesta, interesul respectă mai degrabă pasiunea decât rațiunea, ceea ce sugerează că pasiunea – sau această pasiune în special –este mai presus de interes în ierarhia normativă. Alte pasiuni, precum invidia, se prea poate să fie mai prejos decât interesul. Putem fi martorii comiterii unor acțiuni bazate pe invidie, dar care sunt prezentate în mod plauzibil ca servind unui anumit interes. Nu vor fi întreprinse acțiuni care nu pot fi privite într-o astfel de lumină.

Teoria disonanței cognitive prevede că, atunci când o motivație este *un pic* mai tare decât alta, ea va încerca să recruteze aliați, astfel încât motivele unei tabere să devină în mod decisiv mai puternice. Gândirea inconștientă stă cu ochii în patru, căutând argumente suplimentare în favoarea concluziei provizorii la care a ajuns gândirea

conștientă.[1] În astfel de cazuri, „tăria" motivației nu poate fi considerată ca dată, ci ar trebui văzută, măcar într-o anumită măsură, drept un produs al procesului de luare a unei decizii. Să presupunem că, atunci când cumpăr o mașină, acord anumite valori unor caracteristici cu pondere diferită (viteză, preț, confort, aspect) pentru fiecare dintre opțiuni și ajung la o evaluare globală prin compararea ponderată a sumei valorilor. De exemplu, pot acorda o valoare globală de 50 brandului A și una de 48 brandului B. Din cauza apropierii incomode a termenilor comparați, inconștient voi modifica ponderile, astfel încât A devine un câștigător net cu 60 *versus* 45, ca valoare globală. Înainte să cumpăr produsul, dau peste brandul C, care ar fi înregistrat o valoare de 55 conform vechilor ponderi, dar care atinge doar 50 cu ponderile revizuite. Dacă aș fi văzut opțiunile în ordinea C-A-B, aș fi ales C. Dar, pentru că le-am întâlnit în ordinea A-B-C, aleg A. O asemenea *dependență de cale* subminează ideea simplă conform căreia conflictele motivaționale sunt rezolvate prin apel la tăria motivațională dată.

Conform unei viziuni pe care o putem numi simplistă, decizia asupra furtului unei cărți din bibliotecă poate fi reprezentată în modul următor. Pe de o parte, avem beneficiul utilizării cărții, pe de alta, costul sentimentelor de vinovăție. Rezultatul depinde în bună măsură de stabilirea unui raport de inegalitate între costuri și beneficii. Dar o astfel de perspectivă nu poate fi corectă întrucât putem presupune că mi s-ar oferi o „pastilă a vinovăției", care ar șterge orice sentimente de vină cauzate de furt. Dacă vina intervine în decizie strict ca un cost psihic, utilizarea pastilei ar fi la fel de rațională precum administrarea unei pastile contra mahmurelii pentru cineva care se pregătește de un chef strașnic. Sunt, însă, de părerea că *majoritatea oamenilor s-ar simți la fel de vinovați după administrarea pastilei ca și atunci când ar fi furat cartea.*[2] Nu neg că, într-un anumit mod, ar putea exista un raport de compensare între moralitate și interesul propriu, doar că nu poate fi reprezentat într-un mod atât de simplist.

Un exemplu mai complex este următorul. Îmi doresc să nu-mi fi dorit să nu vreau să mănânc tort. Vreau să mănânc tort pentru că-mi place. Îmi doresc să nu-mi placă pentru că, fiind o persoană moderat vanitoasă, consider că este mult mai important

---

[1] Poate fi întâlnit și fenomenul invers, în care gândirea conștientă încearcă să găsească un motiv pentru o decizie la care a ajuns deja inconștientul. Un exemplu simplu este comportamentul indus de hipnoză. Sau putem lua în considerare explicațiile născocite pe care le găsesc îndrăgostiții pentru a le da telefoane țintei afecțiunii lor. Această formă de autointerpretare nu trebuie confundată cu transmutația, care apare înaintea acțiunii, și nu după.

[2] În capitolul 11 voi argumenta că o persoană cu orizont limitat în timp poate refuza din motive asemănătoare să ia o „pastilă împotriva deprecierii timpului" care l-ar face să acorde o importanță mai mare unor consecințe viitoare ale unor acțiuni curente. Principiul general ilustrat de cele două pastile este că o persoană rațională nu ar vrea să facă în doi pași ceea ce nu ar vrea să facă într-un singur pas. Ar putea, firește, să vrea să facă în doi pași ceea ce nu ar putea să facă într-unul singur.

să-mi păstrez silueta. Dar îmi doresc să fi fost mai puțin vanitos. *Însă este această dorință activată doar atunci când vreau să mănânc tort?* În conflictul dintre dorința mea de a mânca tort, dorința de a fi slab și dorința de a nu fi vanitos, prima și ultima pot forma o alianță și se pot năpusti (sau furișa) peste cea de-a doua. Dacă sunt luat prin surprindere, acestea pot chiar reuși, dar, dacă voi *înțelege* că preeminența dorinței mele de a nu fi vanitos este cauzată de dorința pentru tort, atunci le pot rezista. Cu altă ocazie, dorința mea de gratificare pe termen scurt, precum și dorința mea de spontaneitate pe termen lung pot forma o alianță împotriva dorinței mele de autocontrol pe termen mediu. Atunci când mai mult de două motive au impact asupra alegerii dintre două opțiuni, ideea „puterii motivației" poate fi nedeterminată până când vom ști ce alianță se va forma.

La Bruyère, un moralist francez din secolul al XVII-lea, a rezumat două forme de conflict motivațional în felul următor: „Nimic nu este mai ușor pentru pasiune decât să învingă rațiunea; cel mai mare triumf al pasiunii este, însă, acela de a cuceri interesul." Am remarcat că, atunci când pasiunea „învinge rațiunea", tot s-ar putea să vrea să aibă rațiunea de partea sa. Sfântul Pavel spunea: „Nu fac binele pe care îl vreau, ci răul pe care nu-l vreau", totuși, o reacție mai obișnuită ar fi aceea de a te convinge de binele sau dreptatea a ceea ce vrei să faci în vâltoarea pasiunii. Atunci când pasiunea „cucerește interesul", aceasta se poate întâmpla în mai multe moduri. Agentul poate să nu își ia răgazul să descopere unde anume se află interesul său din cauza *caracterului imperios*, propriu emoției (capitolul 8). De asemenea, forța emoției poate fi atât de puternică, încât aceasta acționează *conștient* împotriva propriului său interes. Un astfel de comportament poate fi echivalent cu slăbiciunea voinței (capitolul 6).

◈◈◈

Notă bibliografică

O teorie a motivațiilor viscerale se poate găsi în G. Loewenstein, „Out of control: Visceral influences on behavior", *Organizational Behavior and Human Decision Processes*, 65 (1996), pp. 272-292. Estimarea excesului de accidente de mașină după 11 septembrie 2001 e luată din G. Gigerenzer, „Dread risk, September 11, and fatal traffic accidents", *Psychological Science*, 15 (2004), pp. 286-287. Lipsa unui astfel de exces în cazul Spaniei este documentat în A. López-Rousseau, „Avoiding the death risk of avoiding a dread risk: The aftermath of March 11 in Spain", *Psychological Science*, 16 (2005), pp. 426-428. Trihotomia interes-rațiune-pasiune este analizată în A. Hirschman, *The Passions and the Interests* (Princeton, NJ, Princeton University Press, 1977), în M. White, *Philosophy, The Federalist, and the Constitution* (Oxford University Press, 1987), precum și în lucrarea mea *Alchemies of the Mind* (Cambridge

University Press, 1999). Mecanismele lipsă corespunzătoare intuițiilor lui Freud provin de la George Ainslie, din lucrarea *Picoeconomics* (Cambridge University Press, 1992). Un studiu clasic al ideii de presiune *versus* atracție îl reprezintă lucrarea lui D. Gambetta, *Did They Jump or Were They Pushed?* (Cambridge University Press, 1983). Am preluat argumentele lui Dion Hrisostomos din P. Veyne, *L'empire gréco-romain* (Paris, Seuil, 2005), p. 217. Principiul din etica iudaică la care am făcut referire este cercetat în D. Daube, *Collaboration with Tyranny in Rabbinic Law* (Oxford University Press, 1965) și în D. Daube, *Appeasement or Resistance* (Berkley, University of California Press, 1987). Am dezvoltat ideea unor stări care sunt esențialmente produse derivate în capitolul 2 din *Sour Grapes* (Cambridge University Press, 1983) și am aplicat-o problemei mântuirii în „Redemption for wrongdoing", *Journal of Conflict Resolution*, 50 (2006), pp. 324-338, iar problema izbăvirii în „Motivations and beliefs in suicide missions", în D. Gambetta (ed.), *Making Sense of Suicide Missions* (Oxford University Press, 2005). Vezi, de asemenea, și L. Ross și R. Nisbett, *The Person and the Situation* (Philadelphia, Temple University Press, 1991), pp. 230-232. Am discutat „deferența față de rațiune" în cazul proceselor colaboraționiștilor după al Doilea Război Mondial în capitolul 8 din *Closing the Books* (Cambridge University Press, 2004). Dovezi pentru modificările în ponderea atașată valorii diferitelor caracteristici ale unor opțiuni alternative se pot găsi în A. Brownstein, „Biased predecision processing", *Psychological Bulletin*, 129 (2003), pp. 545-568, și J. Brehm, „Postdecision changes in the desirability of alternatives", *Journal of Abnormal and Social Psychology*, 52 (1956), pp. 384-389.

B.C.U. „M. EMINESCU" IAȘI

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

Capitolul 5

INTERESUL PERSONAL ȘI ALTRUISMUL

Motivație și comportament

Contrastul dintre motivațiile altruiste și cele ale interesului personal pare simplu la prima vedere. Ca o primă aproximație, vom înțelege prin *motivație altruistă* dorința de a spori bunăstarea celorlalți chiar cu costul unei pierderi a bunăstării personale, iar prin *act altruist*, acțiunea pentru care o motivație altruistă furnizează rațiunea suficientă. Dacă eu văd pe cineva dând bani unui cerșetor pe stradă, numesc acest lucru un act altruist întrucât este o acțiune care *ar putea* să-și aibă originea într-o motivație altruistă, chiar dacă este adevărat sau nu.

Pentru un exemplu mai complex, putem lua în considerare descoperirile experimentale, discutate în mai multe ocazii, legate de „pedeapsa altruistă". În aceste studii, un subiect A are opțiunea de a pedepsi un alt subiect B pentru comportament necooperant, suportând el însuși un cost. Nu există niciun fel de interacțiune față în față și cei doi subiecți nu se vor mai întâlni vreodată. Cu toate acestea, mulți subiecți folosesc opțiunea pedepsei, cauzând un comportament mai cooperant al lui B în interacțiunile ulterioare ale acestuia cu un terț, C. Pedeapsa ar putea să își aibă originea în motivații altruiste dacă A anticipează și este motivat de faptul că pedeapsa aplicată lui B îi aduce un beneficiu lui C. În realitate, este mult mai probabil să fie vorba despre o motivare având-și originea în dorința de răzbunare.

Există multe instanțe ale unui astfel de comportament în afara laboratorului. În Franța secolului al XVIII-lea, exista obiceiul ca țăranii să ofere o masă caldă și adăpost vagabonzilor și cerșetorilor. Dacă un țăran refuza, risca să-și vadă pomii doborâți la pământ, animalele mutilate și casa în flăcări, acte de distrugere care nu produceau niciun beneficiu cerșetorilor și implicau și riscul de a fi prinși. Cu toate că nu avem vreun motiv să credem că motivația lor era, de fapt, aceea de a-l face pe țăran să fie mai binevoitor față de cerșetori în viitor, o astfel de motivație ar fi suficientă pentru o explicație. Răscoalele țărănești din Anglia preindustrială nu aveau sorți de izbândă în atingerea obiectivelor imediate, iar liderii acestora erau de obicei pedepsiți foarte

aspru. Și totuși, în virtutea caracterului lor neplăcut pentru proprietarii de pământuri, răscoalele au generat, pe termen lung, un comportament mai moderat al acestora în raport cu țăranii.

Motivul din spatele definirii motivațiilor altruiste în termeni de sacrificarea bunăstării, mai degrabă decât a bunurilor materiale, rezidă în a exclude cazuri de felul celui următor. Dacă plătesc 100 000 de dolari pentru studiile superioare ale copilului meu, acest lucru se poate datora faptului că bunăstarea copilului meu este atât de puternic legată de a mea, încât „sacrificiul" ne face un bine amândurora.[1] Cu toate că motivația ține cont de celălalt, ea nu este altruistă.[2] Un exemplu de altruism autentic ar fi dacă mi-aș trimite copilul la o școală publică atunci când îmi permit să îl trimit la una privată și să cred că acest lucru este mai bun pentru copilul meu. Făcând aceasta, aș sacrifica nu doar bunăstarea copilului meu, ci și pe a mea. De asemenea, donația la o bancă de sânge (situație diferită de a dona sânge cuiva apropiat) este mult mai probabil să-și aibă originea într-un motiv autentic altruist. Însă, în practică, poate fi imposibil să ne dăm seama dacă o motivație este altruistă sau doar ține cont de celălalt.

Situația se complică și prin faptul că multor oameni le place să dăruiască întrucât se simt bine astfel („efectul căldurii interioare"). Dacă senzația unei călduri interioare este *motivul* pentru care aceste persoane îi ajută pe alții sau fac donații, atunci nu am vrea să îi numim pe aceștia altruiști. Ei dăruiesc pentru că, ținând cont de toți factorii, își sporesc astfel propria bunăstare. Nu vreau să spun că un altruist nu poate simți vreo căldură interioară, ci doar că ea nu intră de obicei în motivele sale (inconștiente) pentru a ajuta sau a dona. Din nou, în practică o astfel de distincție poate fi imposibil de făcut. Voi mai aborda subiectul acesta în capitolul 15.

Oricare ar fi problemele identificării motivațiilor altruiste, avem din plin dovezi ale *comportamentului* altruist. Fundația Carnegie medaliază regulat persoanele care au salvat viața altora expunându-se unui risc personal. Mulți oameni donează sânge fără a primi vreo răsplată.[3] În Norvegia, cele mai multe transplanturi de rinichi se fac prin donații ale rudelor aceluia care primește organul. Extragerea rinichiului presupune un

---

[1] Se poate să fi plătit taxele școlare chiar dacă acestea sunt atât de mari încât costul mi-ar provoca o pierdere în termeni de bunăstare. Într-un astfel de caz, plata unor taxe mai mici este explicată de motive care țin cont de celălalt și care înlocuiesc altruismul.

[2] Când Marx scria că societatea comunistă se va situa dincolo de egoism și altruism și nu că va fi fundamentată de motivații altruiste, e posibil să fi vrut să spună că, orice sacrificii materiale vor fi făcut oamenii unii pentru alții, acest lucru nu implică, din motive asemănătoare, un sacrificiu al bunăstării.

[3] S-a susținut chiar că neplata este importantă pentru a-i elimina pe aceia care au boli infecțioase și care ar vrea să doneze pentru bani.

risc din punct de vedere medical, dar nu aduce nicio răsplată în termeni financiari.[1] Multe persoane, în special femei, au grijă de părinții lor bătrâni pe lângă menținerea unui loc de muncă și grija pentru propriile lor familii. În multe țări, mai mult de jumătate din populația adultă face donații regulate de bani în scopuri caritabile. După tsunamiul din 2004, în multe țări dezvoltate s-au înregistrat vârfuri ale nivelului donațiilor. În vreme de război, unele persoane își ascund anumite handicapuri pentru a li se da voie să lupte. Mulți soldați se oferă voluntari pentru misiuni periculoase (și uneori sinucigașe). Atunci când oamenii votează la alegerile naționale, contribuind în acest mod la viabilitatea democrației, suportă un anumit cost fără vreun beneficiu personal. Lista ar putea continua la nesfârșit.

Motivul pentru care nu putem infera motivațiile altruiste din comportamentul altruist este acela că alte motivații pot *mima altruismul*. În terminologia din capitolul 4, putem vedea altruismul ca o specie de *rațiune*, care poate fi simulată foarte eficient atât de *interes*, cât și de *pasiune*. (Termenul „a imita" sau „a simula" poate să implice, dar nu neapărat, un efort conștient de a induce în eroare cu privire la motivațiile reale.) Mulți oameni care nu se preocupă prea mult de faptul de a fi dezinteresați sunt foarte preocupați de laudele primite pentru caracterul lor dezinteresat. Hume se înșală cu siguranță când susține că „a iubi gloria faptelor virtuoase este o dovadă *sigură* a iubirii de virtute" (italicele îmi aparțin). Dimpotrivă, Montaigne afirmă: „Cu cât este mai strălucitoare fapta, cu atât mai mult îi scad din valoarea sa morală, întrucât se naște în mine suspiciunea că fapta s-a manifestat mai mult pentru strălucirea ei decât pentru bunătate: bunurile puse în vitrină sunt deja pe jumătate vândute." La limită, singurele acte virtuoase sunt acelea care nu ies niciodată la lumină. Bunica angelică a naratorului lui Proust a internalizat acest principiu atât de temeinic, încât își atribuia toate acțiunile bune unor motive egoiste. În măsura în care virtutea are acest caracter discret, e posibil fie mai profundă decât o arată simpla aparență. Firește, din alte motive e posibil să nu fie.

Aprobarea și rușinea

Montaigne a recunoscut și el caracterul rar al virtuții, făcând o distincție între „monedele" motivaționale adevărate și cele false: să acționezi de dragul a ceea ce este drept și să acționezi de dragul a ceea ce cred ceilalți despre tine. Pentru că prima motivație este rară, decidenții trebuie să se bazeze pe cea din urmă:

---

[1] În multe țări, vânzarea de rinichi obținuți de la persoane în viață este ilegală. În acest caz, motivația din spatele legii poate fi aceea de a-i proteja pe nevoiași de ei înșiși, cât și de a-i proteja pe beneficiarii donației de un organ de proastă calitate.

> Dacă acea falsă opinie [preocuparea pentru ce gândesc ceilalți] servește binelui public asigurându-se că oamenii își satisfac îndatoririle, (...) atunci fie ca ea să înflorească nestingherit și să o încurajăm să se dezvolte cât putem de mult. (...) Din moment ce oamenii nu sunt suficient de inteligenți pentru a fi plătiți în monedă autentică, să folosim moneda falsă. Această metodă a fost folosită de toți legislatorii. Și nu există nicio politică pe lume care să nu fi adus anumite onoruri convenționale inutile sau anumite neadevăruri pentru a-i face pe oameni să rămână atașați de datoriile lor.

Cu ocazia instituirii Legiunii de Onoare, în 1802, Napoleon avea o reacție asemănătoare când spunea: „Prin astfel de nimicuri sunt conduși oamenii." (Vechii lui soldați din armata republicană au reacționat vehement împotriva acestei invenții.) *Aprobarea* – dorința de a fi bine văzut de ceilalți – este o monedă falsă, care s-ar putea să se substituie adevăratei monede, a altruismului și moralității. Pe de altă parte, *rușinea* – dorința de a nu fi văzut rău de ceilalți – poate servi ca monedă falsă. Normele sociale îi pot face pe oameni să se abțină de la acțiuni pe care, în alte condiții, le-ar face. Respectarea normelor nu este totuși suficientă pentru a face o impresie plăcută în societate. Aprobarea este rezervată pentru fapte *peste limita obligației*, care trec dincolo de ceea ce impune norma. Ceea ce este obligatoriu într-o societate poate fi peste limita obligației în alta. În Norvegia și în Statele Unite există o normă socială (blândă) conform căreia un frate sau o soră ar trebui să își doneze un rinichi în cazul în care e nevoie și dacă este potrivit pentru transplant[1], pe când în Franța un astfel de comportament poate părea peste limita obligației. În anumite cercuri sociale donațiile sunt obligatorii.

Aceste motivații pot fi ilustrate prin două exemple, aflate într-o puternică opoziție, din politica secolului al XVIII-lea. În prima Adunare Constituantă Franceză (1789-1791), deputații și-au sacrificat în repetate rânduri interese importante, de la renunțarea la privilegiile feudale până la autodeclararea ca neeligibili pentru prima adunare legislativă ordinară. Cu toate că motivațiile lor erau complexe, o componentă importantă era dorința lor de a părea dezinteresați. După cum spunea biograful unuia dintre deputați, trăiau „beția dezinteresului". În aceeași perioadă, în Statele Unite, George Washington și-a manifestat în mod repetat frica de a nu fi considerat de ceilalți prea motivat de interesul personal. (În același timp, era conștient de faptul că prea multă preocupare pentru propria virtute ar fi putut părea nevirtuoasă.) Putem să luăm în considerare două concepții diferite despre onoare. Conform primei concepții, onoarea trebuie *dobândită* prin fapte glorioase. Conform celei de-a doua concepții,

[1] În Statele Unite, doctorii îi ajută de multe ori pe donatorii potențiali să reziste unei astfel de presiuni spunându-le din timp că, dacă li se cere, sunt dispuși să le furnizeze o justificare medicală pentru a nu dona.

onoarea este asumată drept standardul de referință de bază, dar ea poate fi *pierdută* prin fapte rușinoase.

Chestiunea dacă aprobarea sau rușinea pot mima altruismul depinde de criteriile esențiale pe care le aplică ceilalți în evaluarea comportamentului. Unele societăți pot pune un accent puternic pe virtuți care nu tind să imite altruismul într-un mod sistematic, încurajând în acest fel exprimarea unei astfel de virtuți. Dorința pentru onoare poate determina o sumedenie de comportamente care creează risipă din punct de vedere social. Nimicurile lui Napoleon aveau rolul de a-i încuraja pe soldați să își riște viața pentru a spori gloria Franței, nu pentru a crește bunăstarea francezilor. Unii indivizi pot alege o viață trăită în abnegație datorită modului în care societatea îi elogiază pe aceia care duc o viață religioasă virtuoasă, dar pustnicii și călugării sunt cel mai adesea preocupați de ritualurile cultului, și nu de semenii lor. Din câte știu, comunitățile care pun mare preț pe educație și învățare nu au tendința de a genera comportamente mai altruiste decât altele. Cultul frumuseții în societățile occidentale contemporane stimulează comportamente centrate pe sine, care par nefavorabile preocupării pentru ceilalți. În societățile dominate de așa-numitul „amoralism familial" (sudul Italiei a fost dat ca exemplu) există norme sociale împotriva ajutorului acordat străinilor aflați în nevoie sau împotriva legii. Prin urmare, în general, este greu de spus dacă dorința de elogii sau evitarea blamului tind să imite altruismul.

Transmutațiile

Așa cum am discutat în capitolul anterior, motivația originală a unui agent poate fi *transmutată* dinspre interes către rațiune. Mecanismul din spatele unei astfel de alchimii este așa-numitul „amour-propre" sau *amorul propriu*, dorința de stimă și de stimă față de sine. În timp ce aprobarea și rușinea, care sunt derivate din dorința de stimă, afectează doar comportamentul manifest, dorința de stimă de sine poate afecta înseși motivațiile interioare. Cei mai mulți oameni nu vor să se vadă mânați doar de interesul personal. Chiar și atunci când acționează pentru satisfacerea propriilor interese, aceștia încearcă să dea o poleială neegoistă acțiunilor lor. În cele câteva decenii de dinaintea Războiului Civil, sclavia în statele din Sud s-a transformat dintr-o simplă problemă de interes personal într-o adevărată *cauză*, apărată pe temeiuri principiale. Acesta este un tip foarte des întâlnit de ideologie politică. După cum a subliniat și Marx: „Nu trebuie să ne agățăm habotnic de ideea că mica burghezie încearcă în mod principial să impună un interes de clasă egoist. Aceasta vede condițiile *speciale* ale propriei sale emancipări ca fiind aceleași cu condițiile *generale*, care sunt singurele prin care societatea modernă poate fi salvată, iar lupta de clasă evitată."

Cu toate că este greu de spus dacă poleiala neegoistă este sinceră sau ipocrită, ar fi greșit să ne imaginăm că este ipocrită cel mai adesea. De fapt, oamenii au două niveluri de libertate în eforturile lor de a-și justifica propriul comportament. Pe de o parte,

există multe *teorii cauzale cu aspect plauzibil* care pot fi folosite pentru a veni în sprijinul afirmațiilor conform cărora acțiunile benefice pentru sine sunt benefice și pentru ceilalți. Consecința teoriilor creșterii economice „care funcționează de sus în jos", de exemplu, este că un nivel redus al taxelor pentru cei înstăriți va fi benefic pentru săraci. Pe de altă parte, există atât de multe *concepții normative cu aspect plauzibil* ale dreptății, echității sau binelui public, încât ar trebui să fii ghinionist sau incompetent să nu găsești una care (conform unei teorii cauzale cu aspect plauzibil) să nu coincidă cu propriul interes. În capitolul 4 am notat observația că oamenii pot selecta între o filosofie a caritații bazată pe echitate și una bazată pe utilitate pentru a-și justifica donațiile.[1] În astfel de cazuri, oamenii gravitează spontan și inconștient spre o combinație de teorie cauzală și concepție normativă, care poate justifica un comportament aflat pe aceeași lungime de undă cu propriul lor interes. Cu toate că nu înțelegem cum se întâmplă acest lucru, putem fi sigur că se întâmplă.

Nu trebuie ca „mimarea" altruismului să se bazeze pe ipocrizie sau transmutație. Raționarea prin prisma interesului personal poate genera comportamente altruiste în urma unor mecanisme manifeste, cum a fi alegerea *în spatele unui văl al ignoranței* sau reciprocitatea. În capitolul 5 am dat un exemplu al primului mecanism, argumentul lui George Mason conform căruia interesul pe termen lung al familiilor ar trebui să genereze o preocupare pentru bunăstarea tuturor claselor dintr-o societate.[2] Argumente asemănătoare se pot aplica în decursul vieții unui singur individ. În societățile care au o rată mică a șomajului, dar care trec prin schimbări structurale majore, cei mai mulți alegători e posibil să voteze pentru un partid care propune creșterea semnificativă a beneficiilor pentru șomeri dacă aceștia își văd amenințat locul de muncă. Ceea ce părea solidaritate poate fi doar o formă de asigurare.

---

[1] În acest caz, teoria cauzală care justifică donații mici dacă ceilalți dau mult este aceea că donațiile au utilitate marginală în scădere pentru aceia care le primesc. Teoria bazată pe echitate, care justifică donațiile mici dacă ceilalți dau puțin, nu are nevoie de o premisă cauzală.

[2] Guvernatorul Morris a oferit un argument asemănător în dezbaterea privitoare la reprezentarea statelor federale în Senat: „Atașamentul față de State și importanța Statelor a fost un adevărat blestem pentru Țara aceasta. Nu putem nimici șarpele, dar putem, probabil, să îi scoatem colții. El și-a dorit ca ideile noastre să fie extinse pentru a cuprinde adevăratul interes al omului, în loc să fie mărginit la interesul unui anumit Loc. Până la urmă, cât de mărunt poate fi motivul produs de egoism pentru o astfel de politică? Cine poate ști dacă el însuși, cu atât mai puțin copiii săi, vor fi la anul locuitorii unui anumit Stat?" Cu altă ocazie, Guvernatorului Morris însuși i se dădea exact același argument. Ca răspuns la adeziunea lui Elbridge Gerry la propunerea Guvernatorului Morris de a limita reprezentativitatea viitoarelor state din Vest, Roger Sherman a dat următoarea replică: „Trebuie să ne gândim la posteritate, la copiii și la nepoții noștri, care vor fi atât cetățeni ai noilor state din Vest, cât și ai vechilor State. Acest argument singur este suficient pentru a nu pune în practică discriminarea propusă de moțiune."

Reciprocitatea

Reciprocitatea poate fi o relație diadică simplă, ca, de exemplu, în cazul în care fiecare dintre părțile aflate într-o relație este pusă în fața alegerii dintre cooperare și necooperare. Un fermier poate culege recolta în august, altul în septembrie, și fiecare se poate bucura de ajutorul celuilalt. Dacă fermierul a cărui recoltă se coace prima cere ajutorul celuilalt, dar apoi refuză să întoarcă gestul în septembrie, mai mult ca sigur că nu va primi ajutor anul viitor în august. E posibilă dezvoltarea unei relații de asistență mutuală care, cu toate că nu se bazează pe sentimente de camaraderie, le poate cultiva. În timpul Primului Război Mondial, unele trupe germane și britanice au pus la punct un armistițiu tacit, o practică de tipul „trăiește și lasă-l și pe altul să trăiască", bombardând adversarul mult mai puțin agresiv decât ar fi fost în măsură să o facă.[1] Și în acest caz a apărut o atitudine de prietenie față de cealaltă tabără de-a lungul timpului, dar ca un rezultat al cooperării, și nu ca o cauză.

Descartes descrie o formă mai complexă și multilaterală de reciprocitate indirectă:

> Motivul care mă face să cred că aceia care nu fac nimic în afară de urmărirea propriilor foloase ar trebui, de asemenea, dacă vor să fie prudenți, să lucreze, cum fac ceilalți, pentru binele celorlalți și să încerce să-i mulțumească pe toți în măsura posibilităților este că vedem adesea cum se întâmplă că aceia care sunt considerați amabili și prompți în bunăvoință se bucură de o cantitate de fapte bune din partea celorlalți, chiar din partea acelora care nu s-au bucurat la rândul lor de bunăvoință; iar toate acestea nu le-ar primi dacă s-ar considera că au o altă dispoziție; și sacrificiile pe care aceștia le fac pentru binele altor oameni nu sunt la fel de mari precum avantajele aduse de prietenia acelora care le sunt prieteni. Întrucât ceilalți se așteaptă de la noi doar de la fapte pe care le putem face fără un dezavantaj considerabil pentru noi înșine, nici noi nu ne așteptăm la mai mult din partea lor; însă se întâmplă adesea ca faptele care îi costă puțin pe alții să ne fie nouă foarte avantajoase, chiar ne pot salva viața. Este adevărat și că, uneori, facerea de bine e în zadar și că, tot uneori, poți câștiga făcând rău; dar acest lucru nu poate schimba regula prudenței, care este legată de lucrurile care se întâmplă cel mai adesea.

În reciprocitate directă, A îl ajută pe B dacă și numai dacă B l-a ajutat pe A. În reciprocitate indirectă, A îl ajută pe B dacă B l-a ajutat pe C. După cum vom vedea în capitolele următoare, o distincție asemănătoare se aplică și „reciprocității negative": A îl poate prejudicia pe B dacă B l-a prejudiciat pe A, dar și dacă B l-a prejudiciat pe C. Existența reciprocității indirecte sugerează faptul că oamenii se pot comporta altruist pentru a dobândi *reputația* că ar avea motivații altruiste. Ceilalți vor trebui apoi să

[1] Deși Înaltele Comandamente au încercat să oprească această practică, ea s-a dovedit dificil de monitorizat.

decidă dacă acel comportament reflectă un altruism autentic sau este doar o strategie de a dezvolta o anumită reputație de altruism (capitolul 20). În acest caz, reputația este apreciată pe temeiuri instrumentale, și nu intrinseci. În timp ce aprobarea îi cauzează agentului o dorință de stimă de dragul stimei, reputația este căutată pentru beneficiile materiale pe care le poate produce.

Oamenii pot manifesta reciprocitate în situații care nu oferă nicio oportunitate pentru beneficii ulterioare. Dacă A se comportă în mod altruist față de B, B îi poate întoarce serviciul, chiar dacă ambii știu că nu vor mai interacționa vreodată. Fermierul care culege recolta în august îl poate ajuta pe acela care o culege în septembrie chiar dacă plănuiește să emigreze înainte de următorul sezon agricol. Ne putem imagina, firește, motive egoiste pentru reciprocitate. Poate că fermierul care strânge recolta mai devreme se teme de faptul că va fi pedepsit de celălalt în cazul în care nu manifestă reciprocitate sau că va fi marginalizat de terții de ajutorul cărora depinde. Totuși, în condiții de experiment, astfel de efecte sunt excluse. În jocurile experimentale care vor fi discutate mai târziu (capitolele 15 și 20), subiecții interacționează sub protecția anonimatului prin terminale de calculator, eliminând astfel orice fel de efecte ale interacțiunii față în față, cum ar fi rușinea sau jena. De multe ori, jocurile sunt concepute astfel încât o anumită persoană interacționează doar cu un anumit partener.

Chiar și în aceste condiții constrângătoare putem observa reciprocitate. Într-un Joc al Încrederii, un jucător, numit „investitorul", are opțiunea de a transfera unui alt jucător, numit „administrator", oricâte unități, între 0 și 10, din câte are disponibile. Apoi experimentatorul triplează orice sumă trimisă, astfel încât, dacă investitorul trimite 10 unități, administratorul primește 30. Administratorul poate decide să returneze investitorului orice sumă de la 0 la valoarea mărită a sumei (de trei ori valoarea pe care a trimis-o investitorul). Într-unul dintre experimente, investitorii transferă în medie în jur de două treimi din suma disponibilă către administrator, iar administratorii returnează în medie o sumă ușor mai mare. Cu cât este mai mare transferul „către", cu atât va fi mai mare transferul „dinspre". Aceste descoperiri confirmă numeroase presupoziții motivaționale, *cu excepția* aceleia care spune că ambii agenți sunt motivați de interesul personal și știu acest lucru unii despre ceilalți. Într-o astfel de ipoteză, investitorul care așteaptă un transfer zero „dinspre" va face un transfer zero „către". Din moment ce acest rezultat nu a fost observat, înseamnă că sunt în funcțiune motivații care țin cont de celălalt sau de „preferințe sociale". Acestea pot să nu fie suficiente pentru altruism sau echitate din moment ce, în anumite experimente, administratorii nu au returnat nici un ban în plus față de cât au primit, însă chiar și o asemenea sumă tot este mai mare decât ar fi dictat-o interesul personal.

Norme morale, sociale și cvasi-morale

Voi reveni în capitolele următoare la detaliile și consecințele acestui experiment și ale altora înrudite. Aici doar voi face o serie de distincții între trei tipuri de motivații „care țin cont de celălalt". Printre *normele morale* se regăsește aceea de a-i ajuta pe cei aflați în primejdie, norma de a împărți în mod egal și norma „kantianismului de zi cu zi" (fă ceea ce ar fi cel mai bine să faci dacă toată lumea ar face la fel). Printre normele sociale (capitolul 21) se numără normele de etichetă, normele care reglementează răzbunarea și acelea care reglementează utilizarea banilor. Printre așa-numitele „norme cvasi-morale" voi include norma reciprocității (ajută-i pe aceia care te ajută și rănește-i pe aceia care te rănesc) și norma cooperării condiționate (cooperează dacă o fac și ceilalți, dar nu altminteri). Atât normele sociale, cât și cele cvasi-morale sunt condiționate, în sensul că sunt declanșate de prezența sau de comportamentul altora. Susțin că normele sociale sunt declanșate când alții se află în poziția de a observa ceea ce face un agent, iar normele cvasi-morale sunt declanșate când agentul se află în poziția de a vedea ce fac ceilalți.[1] Spre deosebire de acestea, normele morale nu sunt condiționate. Ceea ce ne spun acestea să facem poate fi, cu siguranță, condiționat de ce fac ceilalți. Dacă am o filosofie a carității bazată pe utilitate, cât de mult bine pot să fac (și, deci, cât de mult voi dona) depinde de cât de mult oferă ceilalți. Însă norma în sine nu se referă deloc la alți donatori, ci doar la aceia care primesc.

Voi ilustra distincția dintre norme sociale și norme cvasi-morale făcând apel la două răspunsuri diferite în cazul crizei de apă. În Bogota, în timpul administrației imaginativului primar Antanas Mockus, oamenii urmau o normă cvasi-morală când reduceau consumul de apă. Cu toate că o monitorizare individuală nu era fezabilă, consumul total de apă în oraș era transmis la televizor, astfel încât oamenii își puteau da seama dacă ceilalți se conformează sau nu. Se pare că mulți oameni se conformau pentru a susține cooperarea condiționată. Oamenii își spuneau: „Din moment ce alții consumă mai puțin, e corect să fac și eu la fel." Spre deosebire de acest caz, atunci când apare o criză de apă în California, se pare că operează norme sociale care îi fac pe oameni să își limiteze consumul de apă. Consumul exterior, ca, de exemplu, cel pentru stropirea peluzei, poate fi monitorizat nu doar de vecini, ci chiar de inspectorii municipali. Consumul interior poate fi monitorizat de oaspeți, care își pot exprima dezaprobarea, și chiar o fac, dacă văd că toaleta este curată.[2] De fapt, și în Bogota s-a

---

[1] Cele două se pot susține reciproc când agentul poate să vadă ceea ce fac observatorii înșiși. Dacă te văd că arunci gunoiul pe jos, poate că nu mă deranjează că mă vezi și tu pe mine făcând același lucru. Dar, dacă te văd că pui cu grijă ambalajul în buzunar, echitatea și frica de a fi dezaprobat se pot combina pentru a genera conformare (vezi capitolul 22).

[2] Economia la apă este o problemă și în vremuri normale. În New York se obține printr-o lege care reglează volumul maxim al rezervoarelor de toaletă. În mare parte a Europei se obține

făcut o monitorizare la nivel individual, din moment ce copiii le puteau reproşa părinţilor faptul că nu fac economie la apă.[1]

Normele cvasi-morale pot fi foarte puternice în determinarea unui comportament altruist. Oare acestea *mimează* altruismul sau chiar *sunt* motivaţii altruiste? Motivul pentru care mă refer la ele ca norme cvasi-morale este şi motivul pentru care tind să cred că este vorba despre primul răspuns. Norma reciprocităţii îţi permite să nu îi ajuţi pe cei aflaţi în primejdie decât dacă aceştia te-au ajutat în prealabil. O normă morală tipică este de a-i ajuta necondiţionat pe cei aflaţi în primejdie, chiar dacă nu există antecedente. Norma cooperării condiţionate permite folosirea unei cantităţi normale de apă dacă nimeni altcineva nu reduce consumul, în timp ce atât utilitarismul, cât şi kantianismul de zi cu zi ar sprijini reducerea unilaterală a consumului. Normele morale, am putea spune, sunt *proactive*, în timp ce normele cvasi-morale sunt doar reactive. Un alt mod prin care se poate exprima diferenţa este că sentimentul injustiţiei pare a avea o forţă motivaţională mai puternică decât acela al dreptăţii. După cum vom vedea în capitolul 20, propunerile pe care respondenţii unui experiment tind să le respingă ca nedrepte, cu consecinţa că nici ei, nici propunătorii nu primesc nimic, au acelaşi grad de importanţă cu ceea ce propunătorii tind să ofere atunci când nu sunt constrânşi de frica de a fi respinşi.

Se pare că putem identifica funcţionarea unor motive autentic altruiste dacă sunt îndeplinite două condiţii. Prima este ca acţiunea în beneficiul celorlalţi să fie proactivă, nu reactivă. A doua, să fie anonimă, în sensul că identitatea actorului binefăcător să nu fie cunoscută nici de beneficiar, şi nici de terţi.[2] De exemplu, ne putem imagina o persoană care trimite o sumă de bani în mod anonim ca donaţie la Oxfam sau care pune nişte bani în cutia milei aflându-se într-o biserică pustie. Al doilea exemplu nu este atât de clar pe cât am vrea, din moment ce persoana poate fi motivată de credinţa că Dumnezeu îl observă şi îl va răsplăti. Credinţa poate fi ilogică (o instanţă a „erorii produsului derivat"), dar poate fi destul de răspândită. Primul exemplu poate părea

---

prin instalarea de toalete care au două butoane, fiecare eliberând o anumită cantitate de apă, în funcţie de necesităţi. Ultimul tip de sistem este interesant întrucât nu se bazează nici pe oportunităţi, nici pe stimulente (capitolul 9), ci doar pe bunăvoinţa neobservată a persoanei.

[1] Acest mecanism este sugerat şi de descoperirile experimentale. Într-o campanie de economisire a energiei s-au pus afişe în camerele de duş care le cereau studenţilor să oprească dușul în timp ce se săpunesc şi să-l pornească doar pentru a se clăti. Afişele au avut un efect minim. Însă, imediat ce unul sau doi dintre complicii experimentului au început să se conformeze, numărul celorlalţi studenţi care le-au urmat exemplul a crescut semnificativ. Deşi complicii nu le-au spus nimic celorlalţi, comportamentul lor poate fi considerat drept un reproş tacit adresat acelora care nu s-au conformat.

[2] În experimente, identitatea subiecţilor este ascunsă experimentatorului. În cazul donaţiilor, este ascunsă oficialilor organizaţiei caritabile.

chiar mai ambiguu. Dar până și cele mai pure acte de altruism, precum donațiile anonime făcute unor străini, pot să-și aibă originea în motive de nepătruns. Conform lui Kant,

> este absolut imposibil să observi prin experiență și cu certitudine totală o singură instanță în care maxima unei acțiuni, oricât de dreaptă în sine, se bazează strict pe temeiuri morale și pe concepția despre datorie. Uneori se întâmplă ca, după cea mai atentă autoexaminare, să nu găsim nimic în afara principiului datoriei, care ar fi putut să fie suficient de puternic pentru a ne îmboldi să realizăm cutare sau cutare acțiune făcând sacrificii considerabile; dar nu putem infera cu certitudine din aceasta că nu a fost mai degrabă vorba despre un impuls al dragostei de sine, sub falsă aparență a datoriei, care a fost cauza determinantă a voinței. Ne place să ne flatăm dând credit în mod fals unui motiv mai nobil; pe când, de fapt, nu putem nici chiar prin cea mai severă examinare să scoatem la lumină resorturile secrete ale unei acțiuni, din moment ce, atunci când chestiunea are valoare morală, nu ne preocupă acțiunile pe care le vedem, ci acele principii interne pe care nu le vedem.

Kant spune că, deși nu săvârșim acțiunea în fața unui public exterior, nu putem ști niciodată dacă nu cumva o săvârșim în fața unui *public interior*. Actul de a-ți ascunde virtutea, pe care Montaigne îl găsea atât de virtuos, nu poate fi ascuns față de sine. După cum observa La Rochefoucauld, dragostea de sine „găsește întotdeauna compensații și, chiar și atunci când renunță la vanitate, tot nu pierde nimic". Tot acesta spunea: „Dacă există dragoste curată, liberă de impuritățile celorlalte pasiuni, aceasta este atât de ascunsă în inimile noastre, încât nici noi înșine nu suntem conștienți de ea." În cel mai bun caz, conform lui Proust, putem să ne dăm seama de motivul nostru real făcând apel la ceilalți: „Doar pasiunile celorlalți ne sunt familiare, iar ceea ce ajungem să știm despre propriile noastre pasiuni am fost în măsură să învățăm doar de la alții. Acestea acționează numai indirect asupra noastră, prin imaginația noastră, care înlocuiește motivele noastre originare cu alte motive, care sunt mai acceptabile."

Atribuirea de motivații

Pe lângă motivațiile proprii ale agentului, explicația comportamentului său trebuie adesea să facă apel și la credințele agentului privitoare la motivațiile altora. În formarea acestor credințe agentul înfruntă aceleași dileme hermeneutice cu care se confruntă istoricul sau cercetătorul din științele sociale. Din moment ce agentul nu poate lua în considerare motivațiile declarate ale celorlalți, așa cum sunt declarate acestea, poate în schimb să folosească triangulări generale, de tipul acelora pe care le-am discutat în capitolul 3. În plus, agentul poate folosi tehnici care se aplică doar interacțiunilor față în față. Unii oameni pot fi capabili să identifice un mincinos amator prin

limbajul trupului (sau lipsa acestuia), din moment ce concentrarea acestuia pe ceea ce spune îl face să neglijeze gesturi care nu însoţesc în mod normal vorbirea spontană. De asemenea, pentru a verifica motivaţiile declarate putem întinde o capcană. În timp ce istoricul nu poate să îşi prindă în capcană subiecţii pe care îi studiază, iar cercetătorul din ştiinţele sociale nu poate face de obicei acest lucru din motive etice, angajatorul, partenerul de viaţă sau un părinte se pot simţi mai puţin constrânşi.

Atribuirea de motive celorlalţi este de multe ori afectată de răutate. Putem alege să credem că o acţiune altruistă a fost cauzată de o motivaţie altruistă sau că s-a bazat pe interes personal, însă o presupunem adesea pe cea din urmă, chiar şi atunci când nu avem motive temeinice pentru o astfel de opinie. Cu toate că neîncrederea poate să aibă sens fiind date anumite raţiuni legate de prevedere (capitolul 26), în multe cazuri această justificare nu este disponibilă. Bârfa, de exemplu, pare a fi adeseori motivată de ceea ce moraliştii francezi, urmându-l pe Fericitul Augustin, numeau *caracterul malign* şi *slăbiciunea* naturii umane.[1] Conform lui La Rochefoucauld, „dacă nu am avea defecte, nu ar trebui să ne bucurăm atât de mult văzând defectele altora." De fapt, după cum tot el scrie, dorinţa noastră de a găsi defecte în alţii este atât de puternică, încât ne ajută adesea să le găsim: „Duşmanii noştri sunt mai aproape de adevăr în opinia noastră despre noi decât noi înşine." Dar, cu toate că duşmanii noştri sunt mai aproape de adevăr, şi ei greşesc, chiar dacă mai puţin, însă din direcţia opusă. Pe o scală de la 0 la 10, dacă eu sunt la 6, voi crede că sunt la 9, iar duşmanii mei vor crede că sunt la 4.

Pentru o analiză a acestei atitudini – numită uneori „hermeneutica suspiciunii" – nu pot decât să îl citez pe Jeremy Bentham (am tradus din franceza lui stângace):

> Oricare va fi poziţia asumată de Rege (Ludovic al XVI-lea), oricâte sacrificii va face, nu va reuşi niciodată să-i reducă la tăcere pe calomniatori: aceştia sunt nişte paraziţi pe care o dispoziţie rea şi vanitatea îi vor hrăni întotdeauna, chiar şi în cel mai sănătos regim politic. În primul şi în cel mai important rând, vanitatea este sursa cea mai prolifică a acestei nedreptăţi. Vrem să despicăm firul în patru în orice situaţie (...) şi preferăm cele mai închipuite presupuneri ruşinii de a suspecta că o persoană publică s-ar putea să aibă un motiv lăudabil. Dacă Washington continuă să rămână retras din viaţa politică, preferăm să credem că încearcă să forţeze prin anarhie calea spre despotism. Dacă Necker, în loc să accepte bani pentru serviciile sale, ca oricine altcineva, plăteşte cu propriile sale fonduri pentru a fi lăsat să aducă serviciile respective, preferăm să credem că aceasta nu poate fi decât o stratagemă sofisticată prin care îşi urmăreşte propria lăcomie. Dacă Ludovic al

[1] Nu sunt de acord cu aceia care vor să explice bârfa prin rolul său în constrângerea la a respecta norme sociale. E adevărat că bârfa poate întări sancţiunile informale care susţin normele sociale, dar cred că originea sa are rădăcini mult mai adânci.

XVI-lea renunță la puterea legislativă în favoarea poporului, acest lucru nu poate fi decât rezultatul unui plan elaborat de a o redobândi cu privilegii sporite la momentul oportun.

Ironia face ca ultima acuzație fățarnică citată în textul scris în prima parte a lui 1789 să fie, probabil, justificată în toamna lui 1790. Unul dintre sfetnicii cei mai apropiați ai regelui, Saint-Priest, scrie că, deja la acea dată, regele încetase să se mai opună încălcărilor de prerogative din partea legislativului întrucât „se convinsese de faptul că Adunarea se va discredita prin propriile sale erori". Teoriile conspirației pot fi corecte întrucât chiar există conspirații. Cu toate acestea, descoperirea lor se datorează mai puțin experienței și mai mult unei împotriviri de a recunoaște că persoanele publice pot acționa din motivele corecte.

◈◈◈

Notă bibliografică

Acest capitol se bazează pe lucrarea mea „Altruistic motivations and altruistic behavior", în S.C. Kolm și J.M. Ythier (editori), *Handbook on the Economics of Giving, Reciprocity and Altruism* (Amsterdam, Elsevier, 2006). Alte capitole din acest volum, în special eseul introductiv al lui Kolm, sunt o bună sursă de informații empirice și analize teoretice. Referința la cerșetorii francezi este din G. Lefebvre, *La grande peur* (Paris, Armand Colin, 1988), p. 40, iar cea la răscoalele țărănești din Anglia vine din E.P. Thompson, „The moral economy of the English crowd in the 18th century", *Past and Present*, 80 (1971), pp. 76-136. O analiză a altruismului care ia aspectul „căldurii interioare" se poate găsi în J. Andreoni, „Impure altruism and donations to public goods: A theory of warm-glow giving", *Economic Journal*, 100 (1990), pp. 464-477. Pentru diverse atitudini privitoare la donarea de rinichi, vezi H. Lorenzen și F. Paterson, „Donations from the living: Are the French and Norwegians altruistic?", în J. Elster și N. Herpin (editori), *The Ethics of Medical Choice* (Londra, Pinter, 1994). Am preluat ideea (și termenul) de *approbativeness* („aprobare") din A.O. Lovejoy, *Reflections on Human Nature* (Baltimore, John Hopkins Press, 1961). Rolul dezinteresului în timpul Revoluției Franceze a fost discutat de B.M. Shapiro în „Self-sacrifice, self-interest, or self-defense? The constituent assembly and the 'self-denying ordinance' of May 1791", *French Historical Studies*, 25 (2002), pp. 625-656. Pentru paralela cu cazul american, vezi G. Wood, „Interest and disinterestedness in the making of the constitution", în R. Beeman, S. Botein, și E. Carter II (editori), *Beyond Confederation: Origins of the Constitution and American National Identity* (Chapel Hill, University of North Carolina Press, 1987). Pentru mecanismul transmutației, vezi cartea mea

Alchemies of the Mind (Cambridge University Press, 1999), capitolul 5. Exemplul de tipul „serviciu contra serviciu" din Primul Război Mondial se regăseşte în R. Axelrod, *The Evolution of Cooperation* (New York, Basic Books, 1984). Pentru Jocul Încrederii, vezi C. Camerer, *Behavioral Game Theory* (New York, Russell Sage, 2004), capitolul 2.7. Pentru observaţii privitoare la detectarea minciunilor, vezi P. Ekman, *Telling Lies* (New York, Norton, 1992). Pasajul din Bentham provine din cartea sa *Rights, Representation, and Reform* (Oxford University Press, 2002), pp. 17-18.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 6

MIOPIE ȘI HIPERMETROPIE

Dincolo de urcarea pantei

Principiul plăcerii al lui Freud (capitolul 4) marchează tendința de a căuta satisfacerea imediată a dorințelor. O formă de manifestare a acestei tendințe este adoptarea credinței care se dorește a fi adevărată, și nu pe aceea care este susținută de probe. Iluziile mă fac să mă simt bine aici și acum, chiar dacă, mai târziu, pot să-mi frâng gâtul. Altă formă de manifestare apare în alegerea dintre două acțiuni care determină două fluxuri diferite de utilitate. Principiul plăcerii dictează alegerea acelui flux care are cea mai mare utilitate în prima perioadă, indiferent de forma pe care aceste fluxuri o vor avea mai târziu.

În general, un decident, de la o banală râmă până la o firmă, se poate angaja în *urcarea pantei*. În orice moment în timp, acesta sondează opțiunile *cele mai apropiate* pentru a vedea care dintre ele produce beneficii *imediate* mai mari decât starea de fapt. Limitarea la opțiunile cele mai apropiate este o formă de miopie temporală: principiul plăcerii. Râma sondează mediul înconjurător pentru a vedea dacă vreun loc din jur este mai umed decât cel ocupat în prezent și se va muta în acel loc dacă îl găsește. O firmă sondează „spațiul" celor mai apropiate rutine față de activitatea curentă pentru a găsi una care promite performanțe mai bune pe termen scurt și va adopta o astfel de rutină dacă o găsește. După un timp, râma și firma se vor opri într-un loc superior (pe termen scurt) tuturor pozițiilor din imediata apropiere. S-a atins un *punct de maximum local.*

Ființele umane pot face mai mult. Intenționalitatea – abilitatea de a re-prezenta ceea ce este absent – le permite oamenilor să treacă dincolo de principiul plăcerii și să țină cont de consecințele îndepărtate ale unor alegeri prezente. Planificarea ne permite să facem alegeri care au consecințe mai bune decât acelea care ar decurge din deciziile luate dintr-o clipă în alta sau de la o secundă la alta. În anumite cazuri, oamenii se pot angaja în astfel de acțiuni hipermetropice pentru a satisface mai bine nevoile curente, ca, de exemplu, atunci când un alcoolic renunță să bea un pahar într-un restaurant

BUDAPEST BUSINESS UNIVERSITY LIBRARY

apropiat pentru a putea cumpăra o sticlă întreagă de la un magazin îndepărtat la același preț. În alte cazuri, oamenii se angajează în acțiuni pentru a satisface nevoi viitoare, ca atunci când economisesc pentru bătrânețe. În timp ce primul tip de prevedere este întâlnit și la animale, cel din urmă tip este considerat ca depășind capacitatea acestora. Anumite date recente sugerează, totuși, că și primatele sunt capabile să planifice pe baza unor nevoi așteptate. Oricum, a acționa în baza unor nevoi proiectate este, în mod evident, o operațiune mai sofisticată.

În cele ce urmează, voi da patru exemple de acțiuni bazate pe consecințe îndepărtate din punct de vedere temporal. Primele trei exemple vor fi aduse în discuție și în capitole ulterioare.

RECULER POUR MIEUX SAUTER. Această expresie franceză, echivalentă, în mare, cu „un pas înapoi, doi pași înainte", este ilustrată de faptul fundamental din economie că, în vederea investirii pentru un consum mai ridicat în viitor, trebuie să se consume mai puțin în prezent. Agentul acceptă o stare inferioară stării prezente întrucât ea este o condiție a realizării unei stări alternative superioare mai târziu. Inutil să menționez, acest lucru are sens doar dacă (1) starea inferioară îi permite agentului să supraviețuiască și (2) câștigurile oferite de starea superioară sunt suficient de mari încât să justifice pierderea implicată de instalarea într-o stare inferioară.

AȘTEPTAREA. Multe vinuri, cu toate că sunt foarte bune când sunt îmbuteliate, își sporesc calitățile cu trecerea timpului. Pentru a beneficia de acest fapt, agentul trebuie să fie dispus să respingă o opțiune (să bea vinul imediat), care este superioară stării de fapt întrucât respingerea constituie o condiție a atingerii unui rezultat chiar mai bun mai târziu. Din nou, amânarea consumului poate să nu aibă sens întotdeauna, de exemplu, dacă agentul nu se așteaptă să trăiască suficient de mult încât să se bucure de vinul ale cărui calități au sporit. Pentru un exemplu cu consecințe mai importante, putem să ne gândim la alegerea unui partener de viață. În loc să ceri pe cineva în căsătorie sau să accepți o astfel de cerere cu prima ocazie în care ți se prezintă un partener satisfăcător, mai bine aștepți pe cineva cu mai mult potențial. Riscul, ilustrat din plin în literatura universală, este că s-ar putea să nu dai peste nimeni mai bun.

ESTIMAREA DEPLASĂRII ȚINTEI. Pentru a lovi o țintă în mișcare trebuie ochit nu acolo unde se află obiectul în prezent, ci acolo unde va fi obiectul la momentul întâlnirii cu glonțul. În mod asemănător, ar trebui să țintim în linie dreaptă înspre locul în care se va afla ținta, și nu să urmărim traiectoria curbă indusă de țintirea în poziția curentă.

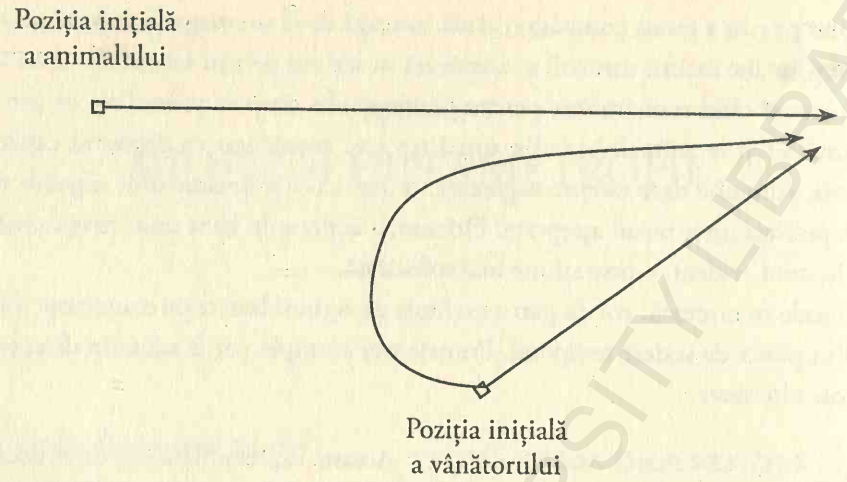

FIGURA 6.1

În Figura 6.1, vânătorul, chiar dacă se mișcă relativ mai încet decât animalul, îl poate prinde mergând în linie dreaptă spre punctul în care animalul se va afla într-un timp calculabil în viitor. Însă, dacă ține întotdeauna în direcția în care se află acum animalul, urmând traiectoria curbă din diagramă, nu îl va vâna niciodată. După cum vom vedea (în capitolul 17), selecția naturală într-un mediu în schimbare poate fi considerată din această perspectivă.

O LINIE DREAPTĂ NU ESTE ÎNTOTDEAUNA CEA MAI RAPIDĂ CALE. Atunci când încerci să atingi o țintă staționară, o linie dreaptă nu este întotdeauna cea mai eficientă cale. În Figura 6.2, salvatorul poate să fugă din instinct înspre înotătorul care se îneacă până ajunge la țărm, iar apoi poate înota distanța rămasă. Dacă ar face o pauză (dar nu prea mare!) pentru a se gândi, și-ar putea da seama că, din moment ce poate ajunge mai repede prin fugă decât înot, ar ajunge la înotător mai repede dacă ar porni pe o cale indirectă, care, cu toate că este mai lungă, se întinde mai puțin în apă. Ne comportăm în modul acesta atunci când mergem pe un drum cu barieră și nu pe drumul care, pe hartă, pare a fi mai scurt. În planificarea economică, un astfel de „comportament de barieră" este de multe ori optim.

Deprecierea timpului

Existența unei capacități de planificare pe termen lung nu presupune și folosirea ei. Pentru ca niște consecințe pe termen lung să schimbe comportamentul prezent, agenții trebuie să fie *motivați* să le ia în considerare. În limbajul psihologilor, aceștia trebuie să fie pregătiți să *amâne satisfacția*. În limbajul economiștilor, aceștia

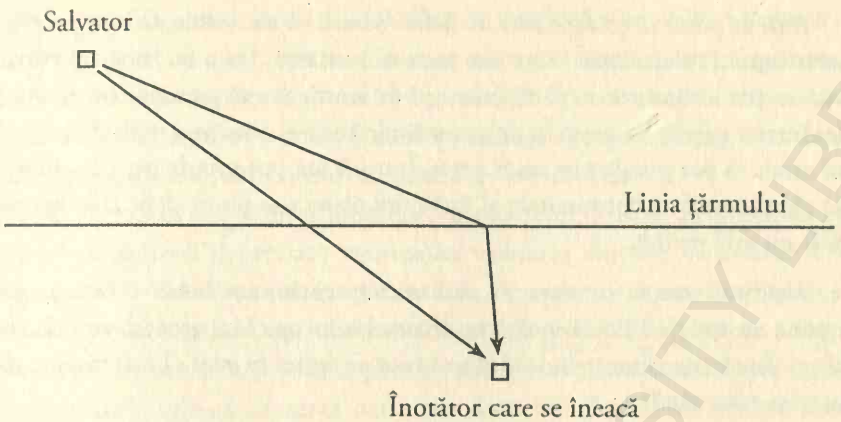

FIGURA 6.2

nu trebuie să fie supuşi unei deprecieri excesive a timpului.[1] Este nevoie de elemente atât motivaţionale, cât şi cognitive. Dacă rezultatele viitoare sunt învăluite în incertitudine, ele nu pot motiva comportamentul prezent. Dacă implică riscul, forţa lor motivaţională este, de asemenea, atenuată. Abilitatea rezultatelor viitoare de a modela comportamentul prezent este afectată atât de *timpul la care*, cât şi de *probabilitatea cu care* vor apărea acestea. Mecanismele prin care acestea afectează alegerea sunt deprecierea timpului şi, respectiv, atitudinea faţă de risc.

Deprecierea timpului (sau miopia) este, după cum o sugerează expresia însăşi, tendinţa de a acorda o importanţă mai mică recompenselor din viitorul îndepărtat faţă de recompensele din viitorul apropiat sau din prezent.[2] Cei mai mulţi oameni puşi să aleagă între 100 de dolari astăzi şi 110 de dolari peste un an ar prefera prima variantă. Această preferinţă poate avea mai multe surse.

Unii oameni pot prefera recompensa mai timpurie întrucât pot investi fondurile respective şi pot obţine mai mult de 110 dolari într-un an.

---

[1] În această carte, expresia „o rată mare a deprecierii timpului" înseamnă că recompensele viitoare au o valoare mică în prezent. Expresia „un factor de depreciere mare" înseamnă că recompensele viitoare au o valoare mare în prezent. Pentru a ilustra şi a motiva această terminologie, aparent ciudată, luaţi în considerare că agentului îi este indiferent dacă va primi 3 unităţi de recompensă mâine şi 2 unităţi astăzi. Recompensa viitoare este redusă (depreciată) cu o treime. Factorul de depreciere a timpului (numărul cu care trebuie să înmulţim recompensa viitoare pentru a obţine valoarea din prezent) este 2/3.

[2] Unii indivizi, precum avarii patologici, pot acorda o importanţă mai mare utilităţii viitoare decât celei prezente. Pentru ei, timpul să fructifici investiţia nu este niciodată potrivit.

Alții ar putea să accepte suta de dolari pentru că au nevoie de bani pentru a supraviețui. Primirea unei sume mai mari de bani mai târziu nu are nicio valoare dacă aceștia se așteaptă ca, până atunci, să fie morți. Sau să presupunem că am de ales între a prinde un pește în râu cu mâinile goale și a încropi o plasă cu ajutorul căreia să pot prinde mai mult pește. Întrucât nu pot prinde pește în timp ce fac plasa, costul de oportunitate al încropirii plasei s-ar putea să fie atât de mare încât nu mi-l permit.

Alții pot accepta recompensa mai mică întrucât sunt bolnavi, ceea ce presupune un risc de 10% să moară în decursul unui an. Mai general, atunci când facem planuri de viitor trebuie să ținem cont de faptul că știm că o să murim, dar nu știm exact când.

Dacă în cazul sumei din viitor avem de a face cu o recompensă așteptată care implică o șansă de 50% de a avea 130 de dolari și o șansă de 50% de a avea 90 de dolari, aversiunea față de risc poate determina o preferință pentru recompensa certă a celor o sută de dolari primiți astăzi.

În ultimul rând, unii oameni poate că preferă recompensa timpurie doar pentru că vine mai devreme. Avem în acest caz o *depreciere pură a timpului*. La fel cum o casă privită din depărtare pare mai mică decât privită de aproape, o sumă mare de bani în viitor poate părea, la nivel subiectiv, mai mică decât o sumă de bani în prezent. Voi lua în considerare doar acest din urmă caz.

Este deprecierea pură a timpului irațională? Să presupunem că o persoană depreciază foarte mult recompensele viitoare. În loc să își asigure o educație superioară, care presupune un sacrificiu temporar al veniturilor pentru un venit mai mare mai târziu, persoana se angajează imediat după terminarea liceului pe un post inferior, care oferă puține posibilități de promovare. Întrucât ignoră impactul pe termen lung al fumatului și al mâncărurilor bogate în colesterol, persoana are o speranță de viață mică.[1] Dacă nu respectă legea pe temeiuri morale, considerațiile legate de prevedere nu vor împiedica o astfel de persoană să le încalce. Este destul de clar că viața acestei persoane va fi scurtă și nefericită. Ce poate fi mai irațional decât așa ceva?

Din punctul meu de vedere, deprecierea pură a timpului nu este irațională în sine. Din cauza acestei deprecieri, viața agentului s-ar putea să o ia pe o cale mai rea decât dacă i-ar fi păsat mai mult de viitor, dar acest lucru poate fi adevărat și cu privire la

[1] Poate că acum cincizeci de ani oamenii ar fi „ignorat" aceste consecințe, adică nu ar fi fost conștienți că ele există. Cu toate că acest lucru este mult mai puțin probabil astăzi, acestea pot fi încă „ignorate", în sensul că li se atașează o importanță mai mică în cadrul deciziei. Destul de des se poate întâmpla ca aceștia să se afle într-o stare de „ignoranță motivată" (o formă de gândire dezideravă) cu privire la consecințe.

motivațiile egoiste. Omul căruia îi pasă doar de propriul sine poate ajunge să aibă o viață tristă și săracă, dar nu trebuie, din acest motiv, să considerăm că egoismul ar fi irațional. Voi discuta aceste chestiuni în capitolul 11. În acest capitol mă voi concentra pe modul potrivit în care poate fi conceptualizată depreciere timpului. Sunt disponibile mai multe abordări, unele cu consecințe radical diferite.

Pentru a modela deprecierea timpului, teoreticienii deciziei au presupus în mod tradițional că oamenii depreciază *exponențial* utilitatea viitoare. O unitate a utilității perioadei t în viitor are valoarea prezentă de k^t, unde $k < 1$ este factorul deprecierii timpului per perioadă. Deprecierea exponențială prezintă un element atractiv din punct de vedere normativ, și anume permite *planificarea consistentă*. Dacă un flux de recompense are o valoare prezentă mai mare decât un altul la un anumit moment, atunci va avea o valoare prezentă mai mare în toate celelalte momente. Prin urmare, agentul nu este supus unei inversări a preferințelor, care este considerată de obicei (în lipsa unor motive pentru răzgândire) drept un punct de referință pentru iraționalitate.

Însă, din punct de vedere empiric, noțiunea de planificare consistentă este mai puțin inteligibilă. Chiar observațiile nesistematice arată un lucru confirmat de observațiile sistematice, potrivit cărora cei mai mulți oameni sunt frecvent supuși unei inversări a preferințelor. De cele mai multe ori, eșuăm în intențiile noastre de a economisi, de a face exerciții fizice dimineața, de a exersa la pian, de a onora o întâlnire. Îmi pot suna dentistul pe 1 martie pentru a mă programa pe 1 aprilie doar pentru a suna pe 30 martie ca să anulez programarea spunând că (din păcate) trebuie să fiu prezent la o înmormântare. Pentru a da seamă de aceste tipuri cotidiene de iraționalitate (și pentru un număr mare de alte fenomene) putem înlocui presupoziția deprecierii exponențiale cu aceea a *deprecierii hiperbolice*.

Să presupunem că valoarea prezentă depreciată a unei unități a utilității perioadei t în viitor este egală cu $1/(1 + kt)$. (În exemplul de mai jos, presupun $k = 1$, dar, în cazul general, k poate fi orice număr mai mare decât zero: cu cât este mai mare, cu atât agentului îi pasă mai puțin de viitor.) Să presupunem, de asemenea, că agentul la $t = 0$, este pus în fața alegerii dintre o recompensă de 10 la $t = 5$ și o recompensă de 30 la $t = 10$. La $t = 0$ valoarea prezentă a celei dintâi este 1,67, iar a celei din urmă este 2,73. Un agent care maximizează valoarea prezentă își va exprima intenția de a alege recompensa amânată. La $t = 1$, valoarea prezentă a recompensei care vine mai devreme este 2, iar a aceleia care vine mai târziu este 3. La $t = 2$ valorile sunt 2,5 și 3,3; la $t = 3$ ele sunt 3,3 și 3,75; la $t = 4$ valorile sunt 5 și 4,29. La un anumit moment, între $t = 3$ și $t = 4$, recompensa care vine mai devreme încetează să fie mai mică și devine opțiunea preferată *doar ca simplu rezultat al trecerii timpului*. Este simplu de remarcat că schimbarea se produce la $t = 3{,}5$, acesta fiind momentul în care dau telefon dentistului pentru a anula programarea.

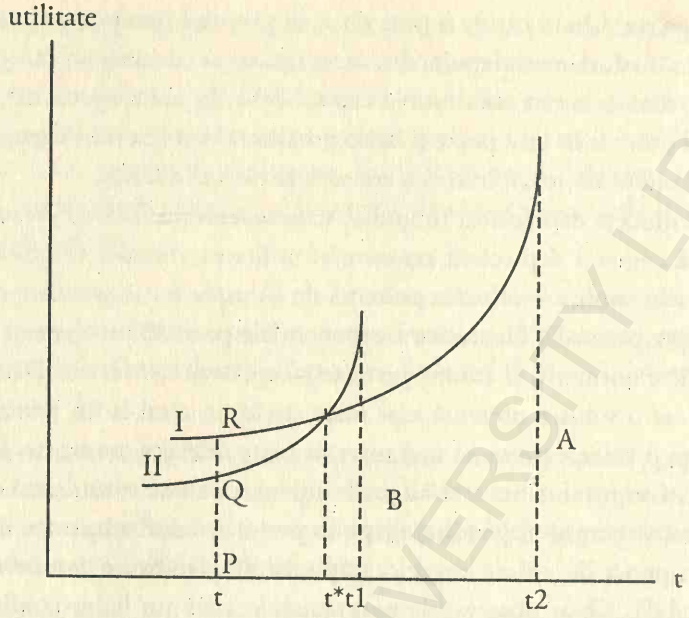

FIGURA 6.3

Diagrama scoate lucrurile în evidență chiar mai bine. În Figura 6.3, agentul poate alege fie recompensa mică B la t_1, fie poate aștepta până la t_2, când ar primi recompensa mai mare A. Curbele hiperbolice I și II reprezintă valorile prezente ale acestor recompense așa cum au fost evaluate mai devreme, în diferite perioade de timp. Acestea sunt, de fapt, *curbe de indiferență* (capitolul 9), care reprezintă compensațiile dintre timpul la care o recompensă devine disponibilă și dimensiunea recompensei. La momentul *t*, de exemplu, agentul este indiferent față de recompensa imediată PQ și recompensa mică de la t_1, așa cum este indiferent față de recompensa imediată PR și recompensa mare la t_2. Din moment ce la momentul *t* valoarea prezentă a lui A este mai mare decât cea a lui B, intenția pe care și-o formează agentul este de a alege A. Dar, întrucât curbele hiperbolice se intersectează la t^*, preferințele se inversează la acel moment, iar agentul alege B.[1]

[1] Există o alternativă, ușor diferită, de a reprezenta deprecierea hiperbolică. Aceasta se bazează pe ideea intuitivă că oamenii fac o distincție radicală între prezent și orice alt moment în timp, acordând o mai mare importanță bunăstării curente decât unei bunăstări în orice alt moment. În plus, oamenii fac distincții *între* perioadele de mai târziu. Într-un exemplu care implică trei perioade de timp, să notăm u_i bunăstarea trăită în perioada *i*, valoarea prezentă sau suma depreciată a utilității fiind $u_1 + b(du_2 + d_2u_3)$. Sunt implicați doi factori de depreciere. În comparație cu prezentul, toată utilitatea viitoare, indiferent de momentul la care este trăită, este depreciată printr-un factor *b*. În plus, oricare dintre utilitățile viitoare sunt depreciate

Pariul lui Pascal

Putem recurge la pariul lui Pascal pentru a pune în evidență relația dintre deprecierea exponențială și cea hiperbolică a timpului. Pascal dorea să-i convingă pe liber-cugetătorii din anturajul său că ar fi mai bine ca aceștia să parieze pe existența lui Dumnezeu din moment ce până și cea mai mică șansă a unei beatitudini eterne este mai atractivă decât cea mai mare dintre plăcerile lumești. Argumentul lui Pascal se evidențiază prin caracteristici foarte complexe, ne vom ocupa de o parte dintre ele în capitolul următor. Pentru moment, vreau doar să atrag atenția asupra unei probleme pe care Pascal nu o menționează: valoarea prezentă (depreciată) a beatitudinii eterne are o valoare finită sau infinită? Dacă este finită, pariorul poate prefera să se bucure de plăcerile lumești decât să aștepte viața de apoi.

Să presupunem, pentru simplificare, că fiecare perioadă din viața de apoi furnizează o unitate de utilitate trăită, că persoana se așteaptă să moară în n ani de la momentul prezent și, de asemenea, că persoana depreciază bunăstarea viitoare exponențial printr-un factor k ($0 < k < 1$). Dacă Dumnezeu există și conferă mântuire persoanei în virtutea credinței acesteia, atunci valoarea prezentă a beatitudinii în primul an după moarte este k^n unități de utilitate, cea din al doilea an este k^{n+1}, și așa mai departe. Aplicând o regulă algebrică elementară, această sumă infinită ($k^n + k^{n+1} + k^{n+2}\ldots$) este egală cu o sumă finită $k^n/(1 - k)$. Cel puțin putem presupune că această sumă poate fi inferioară valorii prezente a n ani de trai hedonist pe Pământ. Dimpotrivă, dacă agentul este supus deprecierii hiperbolice, atunci suma infinită $1/(n + 1) + 1/(n + 2) + 1/(n + 3)\ldots$ crește dincolo de orice valoare finită, ceea ce înseamnă că, prin compararea valorilor prezente, orice plăcere lumească va fi depășită, în cele din urmă, de beatitudinea mântuirii. Chiar dacă cea din urmă este înmulțită cu o probabilitate oricât de mică a existenței lui Dumnezeu, produsul va fi mai mare decât orice număr finit.

Dar să presupunem că interlocutorul lui Pascal este expus în mod regulat unor oportunități de a paria. Privind în perspectivă, acesta preferă să participe la slujbă mai degrabă decât să parieze întrucât astfel va ajunge să creadă în viața de apoi iar acest lucru îi va da siguranța că se poate aștepta la beatitudine infinită. Însă, prin logica

---

exponențial printr-un factor d. Momentul actual are o preeminență viscerală, care îl face să fie mai important în comparație cu oricare altul, în timp ce perioadele de mai târziu își pierd gradual din puterea de a motiva prin ceva destul de asemănător cu o iluzie optică. Acest model, numit „depreciere cvasi-hiperbolică", are în comun cu deprecierea hiperbolică propriu-zisă faptul că poate determina inversări ale preferințelor. Este diferită prin aceea că valoarea prezentă a unui flux infinit de recompense (precum în pariul lui Pascal) are o sumă finită. Există anumite date din neurofiziologie care confirmă faptul că deprecierea cvasi-hiperbolică, deși este utilizată doar ca aproximare utilă a deprecierii hiperbolice propriu-zise, constituie o reprezentare mult mai exactă a ceea ce se întâmplă.

deprecierii hiperbolice, iminența oportunității de a paria va determina o inversare a preferințelor. Acesta își va forma intenția de a paria doar o singură dată înainte de a participa la slujbă. Va spune precum Fericitul Augustin: „Dă-mi castitate și înfrânare, dar nu încă." Săptămâna următoare se va întâmpla același lucru. Astfel, structura însăși a deprecierii timpului care asigură o valoare prezentă mai mare beatitudinii eterne îl va împiedica pe parior să facă primii pași spre a o atinge.

Slăbiciunea voinței

După cum ilustrează acest exemplu, deprecierea hiperbolică poate pune în lumină problema clasică a slăbiciunii voinței (SV). O persoană cu voința slabă (sau un *acratic*) are următoarele caracteristici:

1. Persoana are un motiv pentru a face X.
2. Persoana are un motiv pentru a face Y.
3. Conform propriei judecăți a persoanei, motivul pentru a face X are o greutate mai mare decât cel pentru a face Y.
4. Persoana face Y.

În particular, se consideră că emoțiile au capacitatea de a determina acțiuni împotriva a ceea ce agentul crede că e cel mai bine. Când Medeea din piesa lui Euripide este pe cale de a-și ucide copiii, aceasta spune: „Știu, cu adevărat, ce rău am de gând să săvârșesc. Dar este mai puternică mânia mea decât orice alte gânduri." În versiunea dată de Ovidiu piesei, aceasta spune: „Un impuls necunoscut mă trage-n jos împotriva mea. Mânată încolo și-ncoace, (...) văd binele și îl încuviințez, însă urmez doar răul."

Aceste replici, ca și cele patru enunțuri folosite pentru a caracteriza slăbiciunea voinței, sunt ambigue și insuficient delimitate, în sensul că nu se menționează *când* anume sunt adevărate. Să definim *o concepție strictă a slăbiciunii voinței*, după cum urmează:

1. Persoana are un motiv pentru a face X.
2. Persoana are un motiv pentru a face Y.
3. Persoana face Y, gândindu-se, *la momentul acțiunii*, că motivul pentru a face X are mai multă greutate decât motivul pentru a face Y.

Să ne imaginăm o persoană care s-a decis să se lase de fumat și merge la o petrecere unde i se oferă o țigară. Aceasta acceptă oferta, știind, în timp ce o acceptă, că nu ar trebui să fumeze. O persoană care ține regim poate accepta oferta unui desert știind, în timp ce acceptă, că nu este o idee prea bună. Deși nu este nimic imposibil

în legătură cu această concepție asupra slăbiciunii voinței, intervin anumite probleme empirice. Ar fi destul de greu de stabilit că acțiunea și „gândirea a ceea ce este mai bine" coexistă chiar în același moment în timp, ci, mai degrabă, că gândirea se modifică cu o fracțiune de secundă înainte de comiterea acțiunii. De asemenea, din câte știu, nimeni nu a specificat mecanismul cauzal prin care dorința de a face Y dobândește o mai mare eficiență cauzală decât dorința de a face X.

Pentru a ocoli aceste probleme, putem stabili o *concepție cuprinzătoare asupra slăbiciunii voinței*, care permite gândirii agentului să înțeleagă că faptul de a trebui să facă X și alegerea lui Y se petrec în momente distincte:

1. Persoana are un motiv să facă X.
2. Persoana are un motiv să facă Y.
3. Când persoana se gândește la ce e mai bine pentru sine într-un mod calm, motivul pentru a face X este mai puternic decât cel pentru a face Y.
4. Persoana face Y.

Socrate a negat faptul că SV este posibilă în sens strict. Și Aristotel se apropie de o sugestie similară. Cel din urmă îngăduie un sens cuprinzător al SV dând ca exemplu o persoană a cărei gândire, la momentul acțiunii, se află sub influența alcoolului. Să presupunem că merg la o petrecere cu colegii de serviciu, că beau prea multe pahare de alcool, că îmi jignesc șeful și îi fac avansuri soției sale. La momentul respectiv, toate acestea par cât se poate de firești. Însă dacă, în prealabil, cineva mi-ar fi sugerat că mă voi comporta în acest mod, aș fi respins o asemenea sugestie drept inconsistentă cu gândirea mea calmă. Dacă aș fi fost convins de faptul că gândirea mea se va îneca în alcool, aș fi stat deoparte. După cele întâmplate, poate că îmi regret amarnic comportamentul.

Acest caz, ilustrat de Figura 6.4, este un caz al *inversării temporare a preferinței*, și nu al SV propriu-zise. Există cel puțin trei mecanisme care pot determina aceste schimbări. Primul este *proximitatea temporală*, așa cum este explicată în discuția deprecierii hiperbolice. Al doilea este *proximitatea spațială*, ilustrată de fenomenul dependenței de ocazie. Acest mecanism explică, de exemplu, cazurile de recidivă a dependenților. Chiar după ani de abstinență, o ocazie prezentată de mediu asociată în mod tradițional cu consumul de droguri poate declanșa recidiva. Poate fi de ajuns să vezi la televizor câteva obiecte asociate cu drogurile. Decizia de a urma un regim poate

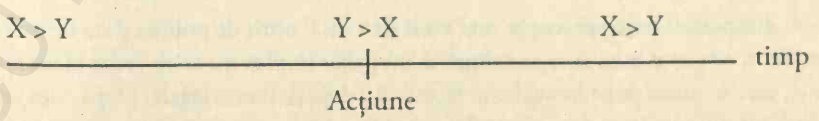

FIGURA 6.4

BCU IASI CENTRAL UNIVERSITY LIBRARY

fi subminată de apariția unei mașini care transportă dulciuri. Și în aceste cazuri agentul alege în conformitate cu concepția sa la momentul alegerii privitoare la ceea ce preferă mai mult, toate lucrurile fiind luate în considerare. În cele din urmă, *pasiunile* au capacitatea de a determina o schimbare temporară de preferințe prin faptul că acestea au un timp de înjumătățire foarte scurt (capitolul 8). Ele pot determina o inversare a preferințelor, cauzând o scădere a atenției acordată de agent viitorului îndepărtat.[1]

Putem extinde această idee pentru a include și schimbările temporare (și motivate) ale *credințelor* agentului. Într-o astfel de concepție cuprinzătoare, slăbiciunea voinței poate fi provocată de autoînșelare (sau de iluzie). Cu toate că, în prealabil, persoana respectivă a decis să nu consume mai mult de două pahare pentru a putea fi în stare să conducă acasă în siguranță, sub influența dorinței pentru un al treilea pahar aceasta poate să-și spună că încă un pahar nu are cum să îi influențeze atât de mult abilitatea de a conduce.[2] Preferința sa (pentru conducerea în siguranță) rămâne neschimbată, dar credința referitoare la condițiile în care poate conduce în siguranță s-a modificat. De asemenea, persoana poate suferi o modificare temporară a preferințelor dacă se decide că distracția la petrecere este atât de importantă încât merită să își asume riscurile (pe care le poate percepe foarte precis) ale conducerii sub influența alcoolului.

◈◈◈

Notă bibliografică

Pentru informații privitoare la faptul că primatele pot fi în măsură să estimeze nevoi viitoare (care nu sunt experimentate curent), vezi N. Mulcahy și J. Call, „Apes save tools for future use", *Science*, 312 (2006), pp. 1038-1040. Două cărți-sursă

---

[1] De fapt, inversarea preferințelor cauzată de preferința de timp hiperbolică poate fi imitată de schimbările induse de emoție asupra factorului de depreciere asociat cu preferințele de timp exponențiale. Să presupunem că agentul are de ales între două opțiuni, A și B, care oferă recompensele (2, 5, 6) și, respectiv, (5, 4, 1) în trei perioade de timp succesive. Cu o rată a deprecierii pe perioadă de 0,8 (și o rată pe două perioade de 0,64), valorile prezente ale celor două opțiuni (așa cum sunt acestea evaluate în prima perioadă) sunt 9,84 și, respectiv, 8,84. Cu o rată de depreciere pe perioadă de 0,6 (și o rată pe două perioade de 0,36), valorile sunt 7,16 și 7,96. Nu este de mirare că agentul încetează să prefere opțiunea cu consecințe mai bune pe termen lung când emoțiile îl fac să acorde o atenție mai mică viitorului.

[2] Dacă, dimpotrivă, este preocupat mai mult de a nu fi oprit de poliție, și nu de perspectiva unui accident, este mai greu să se convingă că un pahar în plus nu se va vedea la un test al alcoolemiei, care ar putea pune în evidență faptul că a depășit limita legală. După cum voi argumenta în capitolul următor, chiar și gândirea deziderativă este supusă (într-o oarecare măsură) unor constrângeri ale realității.

pentru deprecierea timpului și alte aspecte ale alegerii intertemporale aparțin lui G. Loewenstein și J. Elster (editori), *Choice over Time* (New York, Russell Sage Foundation, 1992) și G. Loewenstein, D. Read și R. Baumeister (editori), *Time and Decision* (New York, Russell Sage Foundation, 2003). Am discutat pariul lui Pascal *in extenso* în „Pascal and decision theory", în N. Hammond (ed.), *The Cambridge Companion to Pascal* (Cambridge University Press, 2004). Dovezi venite dinspre neurofiziologie care pun în evidență deprecierea cvasi-hiperbolică a timpului pot fi găsite în S. McClure *et al.*, „Separate neural systems evaluate immediate and delayed monetary rewards", *Science*, 306 (2004), pp. 503-507. Discuțiile moderne asupra slăbiciunii voinței au fost relansate de D. Davidson în „How is weakness of the will possible?", în cartea sa *Essays on Action and Events* (Oxford University Press, 1980). Am comentat ideile sale în „Davidson on weakness of will and self-deception", în L. Hahn (ed.), *The Philosophy of Donald Davidson* (Chicago, Open Court, 1999). Formarea motivată de credințe este supusă unei discuții în D. Pears, *Motivated Irrationality* (Oxford University Press, 1984). Am discutat pe larg legătura dintre slăbiciunea voinței și inversarea preferințelor în „Weakness of will and preference reversal", în J. Elster *et al.* (ed.), *Understanding Choice, Explaining Behavior: Essays in Honor of Ole-Jørgen Skog* (Oslo Academic Press, 2006).

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 7
CREDINȚELE

Ce înseamnă a „crede" ceva?

Pentru a înțelege rolul credințelor în generarea acțiunii, trebuie să le înțelegem natura, cauzele și consecințele. După cum am precizat în observațiile introductive făcute la Partea a II-a, nu este întotdeauna clar ce înseamnă a „crede" că un lucru există, de exemplu, că există viață după moarte. Multe dintre personalitățile religioase de prim rang au scris despre lupta constantă cu îndoiala. Oare în momentele în care aceștia credeau, credința lor era la fel de simplă și necondiționată precum aceea a cuiva care nu cunoaște îndoiala? Oare adepții comunismului care „credeau" că Partidul nu poate să greșească chiar *credeau* acest lucru?[1] Cum ne putem da seama de diferența dintre pesimistul congenital, care are tendința să creadă tot ce e mai rău, și decidentul prudent, care acționează *ca și cum* s-ar putea adeveri cel mai rău scenariu?

De asemenea, în limbajul obișnuit „a crede" este un termen care sugerează ceva mai puțin decât atașamentul total. *Cred* că mâine o să plouă, dar știu că s-ar putea să mă înșel. Nu pot spune că doar cred că sunt căsătorit din moment ce *știu* asta. În analizele filosofice, cunoașterea este definită de obicei drept opinia (credința) adevărată și întemeiată, o credință care stă într-o relație de un anumit tip atât cu lumea (este adevărată), dar și cu un corp de probe pe care le are agentul (este întemeiată). Și totuși, niciuna dintre aceste trăsături ale cunoașterii nu surprinde certitudinea subiectivă care însoțește atât de des expresia „știu" în discursul de zi cu zi. Această certitudine nu este doar limita de probabilitate de 97%, 98%, 99%, 99,9% și așa mai departe.

---

[1] După căderea comunismului, la o întrunire publică o femeie din fosta Germanie de Est a spus că generația sa fusese învățată din copilărie să se conformeze, să nu iasă în evidență. O schizofrenie extinsă pe o perioadă lungă de timp *i-a golit de umanitate*. Prin urmare, a continuat femeia, nu putea acum să înceapă din senin să „vorbească deschis" sau „să spună ce crede". Ea nici măcar nu știa exact ce anume crede.

BUCHAREST UNIVERSITY LIBRARY

Este diferită din punct de vedere calitativ de orice element mai slab decât certitudinea absolută.[1]

Acest „efect al certitudinii" apare în următorul experiment. Un grup de subiecți a fost rugat să își exprime preferințele pentru mai multe opțiuni. (Numerele dintre paranteze indică proporția subiecților care preferă o anumită opțiune.)

O șansă de 50% de a câștiga o excursie de trei săptămâni în Anglia, Franța și Italia (22%).
O excursie de o săptămână în Anglia, cu certitudine (78%).

Unui alt grup i s-au dat aceste opțiuni:

O șansă de 5% de a câștiga un tur de trei săptămâni în Anglia, Franța și Italia (67%).
O șansă de 10% de a câștiga un tur al Angliei (33%).

Membrii primului grup tind să prefere opțiunea „doar Anglia" pentru că este disponibilă *cu siguranță*. Odată ce este redusă cu aceeași probabilitate ca alternativa, cea din urmă opțiune pare mai atractivă. Soldaților cărora li se cere să se ofere voluntari pentru misiuni foarte periculoase pot avea considerabil mai puține ezitări decât aceia cărora li se cere să se ofere voluntari pentru misiuni sinucigașe. Sigur, în cazul primilor poate funcționa gândirea deziderativă („Nu are cum să mi se întâmple mie"), care nu poate funcționa în cazul celor din urmă.

Patru atitudini cognitive

Chiar dacă lăsăm la o parte aceste probleme, ideea de credință rămâne ambiguă. Putem distinge între patru tipuri de atitudine cognitivă în raport cu lumea, în funcție de puterea lor. Primul mod este *certitudinea*. Al doilea este *riscul*, un mod în care agenții atribuie probabilități, fie bazate pe frecvențe trecute, fie pe propria lor judecată, pentru fiecare posibilitate a unui set de rezultate mutual exclusive și exhaustive luate împreună. Al treilea mod este cel al *incertitudinii*, în care oamenii cunosc setul de rezultate mutual exclusive și exhaustive luate împreună, dar nu sunt capabili să le atribuie vreo probabilitate (cardinală).[2] În cele din urmă, al patrulea mod este *ignoranța*, în care atât arealul de rezultate posibile, cât și probabilitățile acestora

---

[1] Chestiunea este ceva mai complicată. Când oamenii se confruntă cu probabilități mari din punct de vedere numeric, precum 99,9%, se pot declanșa două mecanisme (în sensul capitolului 2): diferența dintre incertitudine și probabilitate mare este neglijată sau exagerată. (Asemănător, probabilitățile foarte mici sunt fie neglijate, fie exagerate.) Această complicație nu apare, însă, în exemplul dat în text referitor la efectul certitudinii.

[2] Aceștia pot fi capabili să atribuie probabilități ordinale, adică să spună că un rezultat este mai probabil să se întâmple decât altul, fără să fie în măsură să spună cât de probabil.

BCU CLUJ CENTRAL UNIVERSITY LIBRARY

sunt complet necunoscute. În cuvintele memorabile ale fostului ministru al apărării Donald Rumsfeld, nu suntem doar puși în fața unor elemente cunoscute și necunoscute, ci și în fața unor „necunoscute necunoscute".[1]

Mă voi concentra pe certitudine și risc, nu pentru că acestea sunt atitudinile cognitive potrivite întotdeauna, ci pentru că sunt cel mai des întâlnite. Chiar atunci când oamenii nu au *niciun* motiv pentru a avea anumite credințe referitoare la un anumit subiect, aceștia se simt îndemnați într-un mod irezistibil în a-și forma o opinie – nu o opinie determinată (cum este cazul gândirii dezideravive), ci o opinie *oarecare*. Această propensiune este determinată de factori culturali într-o anumită măsură. Albert Hirschman susține că cele mai multe culturi din America Latină „pun o valoare deosebită pe ideea de a avea *opinii puternice* în legătură cu aproape *orice* de la bun *început*". În astfel de societăți, faptul de a admite ignoranța echivalează cu faptul de a admite înfrângerea. Dar tendința aceasta este realmente universală. Montaigne spunea că „multe dintre abuzurile lumii acesteia sunt produse – în treacăt fie spus, toate abuzurile lumii acesteia sunt produse – de faptul că suntem învățați să ne fie teamă să admitem că suntem ignoranți și pentru că ni se cere să acceptăm orice nu poate fi respins". Intoleranța față de incertitudine și ignoranță își are originea nu doar în orgoliu, ci și în dorința umană universală de a găsi semnificații și un caracter sistematic peste tot. Gândirea are oroare de vid.

O versiune particulară a tendinței de a găsi semnificații peste tot în univers este de a atribui unor evenimente mai mult sau mai puțin aleatoare capacitatea de a *acționa*. În timpul Vechiului Regim, în Franța, populația nu a fost capabilă să accepte că numai natura este responsabilă pentru nefericirea sa. În timpul unei penurii de grâne, părerea generală era că speculanții care își făcuseră depozite de grâne umflaseră foarte mult prețul, chiar dacă, de fapt, cauza era recolta foarte proastă. Alteori, lipsa grânelor era explicată prin dorința elitelor de a-i înfometa pe oameni, ca parte a unui neîntrerupt război de clasă. Conform Articolului 22 din Carta Mișcării Hamas, evreii au fost „în spatele Revoluției Franceze, al revoluției comuniste, precum și în spatele tuturor mișcărilor revoluționare din trecut și din prezent de oriunde în lume. Cu banii lor, aceștia au format societăți secrete, precum Francmasoneria, cluburile Rotary, Lions și altele, în diferite locuri ale lumii, cu scopul de a sabota respectivele societăți și pentru a satisface interese cu caracter sionist. Cu ajutorul banilor, au reușit să țină sub control statele imperialiste și să le instige să colonizeze cât mai multe state pentru a le putea exploata resursele și pentru a răspândi și acolo corupția." O asemenea stare mentală conspirativă sau paranoică este imună la orice fel de contraargument din moment ce aceia care cred într-o teorie a conspirației vor găsi *confirmarea* ei în lipsa probelor, ba

[1] Sunt posibile cazuri mai complexe. Pot fi în măsură să atribui valori de probabilitate unor rezultate, în timp ce nu sunt în măsură nici măcar să le specific pe celelalte.

chiar și în existența unor probe care indică exact contrariul, lucruri pe care ei le interpretează drept semne ale naturii diabolic de inteligente a conspiratorilor.

Mecanismele generatoare de eroare se bazează cumva pe *motivație*. Însă eroarea poate să își aibă originea și în *ignoranță*. Argumentul pare evident, dar este ceva mai subtil. Darwin, de exemplu, observă că „ignoranța produce încredere mai frecvent decât cunoașterea". Ignoranța, împreună cu încrederea, constituie o rețetă bună pentru eroare. Invers, prin extinderea cercului de lumină, crește și zona întunecoasă care îl înconjoară, ceea ce produce o mai mare umilință. De fapt, experiența sugerează că incompetența nu cauzează doar performanțe cognitive modeste, ci și o inabilitate de a recunoaște cât de mare este incompetența însăși. Incompetenții suferă de un dublu handicap.

Însă mai există o posibilitate, anume aceea că „primejdia" o reprezintă *„puțina* cunoaștere", și nu ignoranța. Montaigne scrie că „există o ignoranță de școală primară, care precedă cunoașterea, și o ignoranță doctorală, care o urmează"; Pascal spune același lucru, chiar mai pe larg. Prin achiziționarea unei cantități mai mari de informație, devii mai întâi foarte încrezător, pentru ca, în final, să fii mai puțin încrezător. Bănuiesc că ipoteza Montaigne-Pascal are mai multă validitate decât aceea a lui Darwin. *Creșterea* nivelului cunoașterii, și nu *nivelul* crescut al cunoașterii, îndeamnă gândirea să tragă concluzii care nu sunt susținute de date din cauza unei încrederi prea mari în forțele proprii.

Evaluări subiective ale probabilității

Judecățile de probabilitate își pot avea originea în observarea frecvențelor obiective sau pot fi evaluări pur subiective.[1] Când agentul se poate baza pe un număr mare de observații ale unor situații asemănătoare, metoda frecventistă dă rezultatele cele mai bune. Dacă plănuiesc să merg la un picnic luna viitoare de ziua mea și trebuie să îmi formez o opinie despre cum va fi vremea, cel mai bun lucru pe care îl pot face este să mă uit la statisticile meteorologice din anii precedenți pentru ziua respectivă. Însă, dacă trebuie să îmi formez o opinie despre cum va fi vremea mâine, cel mai bun factor singular predictiv este vremea de astăzi. Nu este, însă, singurul factor predictiv. Datele înregistrate în trecut îmi pot spune dacă o vreme însorită este un eveniment rar sau normal pentru ziua respectivă. Dacă este un eveniment rar, atunci faptul că azi este frumos afară își pierde din valoarea de factor predictiv. Pot consulta barometrul de pe perete pentru a vedea dacă presiunea atmosferică crește sau scade ori mă pot uita pe cer să urmăresc zborul rândunicilor și așa mai departe.

---

[1] La o analiză mai atentă, prima metodă (obiectivă) se reduce la a doua (subiectivă) din moment ce, pentru ca datele obiective să fie utile, ele au nevoie de o interpretare subiectivă. Însă în practică această distincție este clară și folositoare.

A integra toate aceste informații într-o judecată probabilistă despre cum va fi vremea mâine este o sarcină dificilă. Cei mai mulți dintre noi nu ne pricepem foarte bine să facem așa ceva. De cele mai multe ori, problema nu este lipsa de informație, ci, mai degrabă, abundența de date combinată cu o lipsă a procedurilor formale de integrare a datelor într-o opinie care să țină cont de toți factorii dați. Unii oameni sunt, totuși, mai buni decât cei mai mulți dintre noi în integrarea informațiilor vaste și difuze, cu grade variate de relevanță, într-o evaluare globală. Aceștia posedă calitatea crucială, dar evazivă, a puterii de *judecată*. Generalii de succes, oamenii de afaceri și politicienii tind să aibă această calitate, tocmai mulțumită ei au succes. Un bun președinte al unei bănci centrale are nevoie de această calitate, însă celor mai mulți dintre economiști le lipsește.[1] Cel mai bun lucru care ne rămâne de făcut, nouă celorlalți, este să recunoaștem că nu avem această calitate și să învățăm să nu avem încredere în intuiția noastră. Pot învăța, de exemplu, că, de multe ori, nu am încredere în oameni din motive irelevante, odată ce am ajuns să îi cunosc pe acei oameni. („Arăta ca un golan pe care-l știam într-a cincea.") Prin urmare, pot ajunge să fiu neîncrezător în propria mea neîncredere.[2]

Și totuși, avem tendința să credem că nu doar generalii de succes, politicienii și oamenii de afaceri au judecată, ci și experții cu experiență. În chestiunile complicate de diagnostic sau prognoză – precum identificarea indivizilor bolnavi psihic sau evaluarea probabilității ca o persoană care cere eliberarea condiționată din închisoare să comită o nouă infracțiune – avem încredere în expert. Datorită experienței lor, experții sunt sensibili la semnele grăitoare pe care observatorii neantrenați le pot ignora sau a căror semnificație nu o pot înțelege. Mai mult, când probe diferite indică direcții diferite, experții se pot baza pe experiența lor pentru a decide cărei probe i se va acorda o importanță mai mare. Cel puțin așa ne imaginăm că stau lucrurile cu experții. Întrucât cei mai mulți dintre noi ne considerăm experți într-un domeniu sau altul – măcar când vine vorba să prezicem comportamentul șefului nostru, al partenerului de viață, al copiilor –, am investit suficient de mult în această imagine a abilităților cognitive superioare ale expertului.

Din păcate, această imagine este în totalitate falsă. În multe studii, performanța experților de a oferi un diagnostic sau o prognoză a fost comparată cu performanța

---

[1] Scriind despre Alan Greenspan (*New York Times*, 28 octombrie, 2005), Paul Krugman observă că acesta, în timp ce manifesta neîncredere față de modelele formale, avea „abilitatea de a ghici din date fragmentare și uneori contradictorii încotro bate vântul în economie".

[2] Să știm că suntem supuși greșelii este un lucru; să o putem îndrepta este alt lucru. Studiile arată că încercările deliberate de a elibera de erori propria judecată nu au o valoarea prea mare din moment ce cădem foarte ușor în capcana corecturilor insuficiente, a corecturilor inutile sau a corecturilor exagerate.

unei simple formule mecanice bazată pe câteva variabile. În esență, acest lucru se reduce la faptul de a compara metodele obiective (frecventiste) și metodele subiective. Ponderile atribuite variabilelor sunt derivate prin tehnici statistice care precizează ponderile care prezic rezultatele cel mai probabil să fie observate. Aproape fără excepție, formula funcționează cel puțin la fel de bine ca expertul, de obicei chiar mai bine.[1] Pentru a da doar un exemplu, într-un studiu pentru diagnosticarea disfuncționalității progresive a creierului, bazat pe testarea capacităților intelectuale, o formulă derivată dintr-un set de cazuri și aplicată ulterior unui nou eșantion a identificat corect 83% dintre cazurile noi. Grupuri de medici experimentați și mai puțin experimentați au identificat corect 63%, respectiv 58% dintre cazuri. Mai mult, experții sunt de cele mai multe ori în puternic dezacord unii cu alții. Într-un alt studiu, psihiatri cu foarte multă experiență care au studiat același interviu psihiatric nu au fost în stare să se pună de acord nici cu privire la diagnosticul pacientului, nici cu privire la motivațiile și sentimentele acestuia. Unii psihoterapeuți folosesc răspunsuri la pete de cerneală ambigue ca indicatori pentru diagnosticare. Însă se pare că pacienții sunt cel puțin la fel de ambigui pentru medici precum sunt acele pete de cerneală pentru pacienți.

Erori ale inferenței statistice

Experții, la fel ca neprofesioniștii, greșesc adesea pentru că ignoră principii evidente sau mai puțin evidente ale raționării statistice. Subiecților unui studiu li s-a dat o descriere a unui tânăr cu părul lung și care obișnuia să citească poezie, fiind rugați să spună dacă li se părea mai probabil ca acesta să fie violonist într-o orchestră sau șofer de camion. Cei mai mulți au spus că acesta este cel mai probabil violonist, ignorând astfel o *trăsătură de bază* a datelor disponibile despre cele două grupuri, și anume numărul absolut de indivizi din cele două grupuri. La nivelul întregii țări, există mai mulți șoferi de camion decât violoniști (și, de asemenea, nenumărate variații în rândul șoferilor de camion), astfel încât este mult mai probabil ca tânărul cu înclinații poetice să conducă un camion.

Altă sursă de greșeli în formarea credințelor este *eroarea de selecție*. Pacienții unui centru de dializă sunt de multe ori surprinzător de reticenți să se treacă pe lista de așteptare pentru un transplant de rinichi. Un motiv este acela că ei au văzut că, în cele mai multe cazuri, transplantul de rinichi nu a reușit, pacienții fiind astfel nevoiți să facă din nou dializă. Montaigne se referă la o eroare de acest tip când spune despre Diagoras că, „prezentându-i-se multe jurăminte și portrete votive de la supraviețuitorii unor naufragii și (...) întrebat apoi: «Hei, tu, care crezi că zeii sunt indiferenți la treburile omenești, ce mai ai de zis despre atât de mulți oameni salvați prin grația

[1] Această superioritate se păstrează chiar dacă atribuim ponderi egale tuturor variabilelor!

lor?», ar fi răspuns: «Lucrurile stau astfel: nu sunt înfățișate aici portretele acelora care nu au fost salvați și s-au înecat, iar acestea sunt mult mai numeroase!»" Asemănător, un psihiatru care susține că „niciun molestator de copii nu se poate opri singur" ignoră faptul că, dacă există vreunul care s-a oprit singur, e puțin probabil să îl fi întâlnit.

Responsabilii forțelor aeriene israeliene au făcut o eroare mai puțin evidentă când au evaluat eficiența relativă a recompensei și a pedepsei în antrenamentul piloților. Observând că performanța piloților crește când sunt pedepsiți pentru o performanță slabă, dar nu atunci când sunt recompensați pentru una foarte bună, aceștia au tras concluzia că pedeapsa este mai eficientă. În felul acesta, au ignorat fenomenul cunoscut ca *regresia către medie*. În orice serie de evenimente care sunt pe deplin sau în parte determinate de șansă, există o tendință ca o valoare extremă la un moment dat să fie urmată de o valoare mai puțin extremă la momentul imediat următor. Tații înalți au fii care sunt mai scunzi decât ei, iar performanțele slabe ale piloților sunt urmate de altele mai puțin slabe, independent de recompense sau pedepse. Când atleții care au un sezon excepțional nu reușesc să repete performanța în sezonul următor, fanii și antrenorii spun de multe ori că au fost prea îmbătați de succes, când ceea ce observăm de fapt poate fi doar regresia către medie.

Eroarea jucătorului de ruletă și judecata ei conversă (care nu are nume) ne pun la dispoziție un alt exemplu. Încheierea asigurărilor de cutremur crește dramatic după un cutremur, apoi scade progresiv pe măsură ce amintirea evenimentului se atenuează. Precum pariorii care fac greșeala să creadă că este mult mai probabil să apară roșu dacă a apărut de mai multe ori la rând, aceia care cumpără o astfel de asigurare își formează credințele folosind *euristica disponibilității*. Judecata lor despre probabilitatea unui eveniment este formată de ușurința cu care acesta poate fi rememorat, iar evenimentele recente sunt mult mai accesibile decât cele trecute. Erodarea emoției de-a lungul timpului (capitolul 8) poate fi, de asemenea, un factor. În direcția opusă, oamenii care trăiesc în zone supuse unor inundații frecvente cred de multe ori că o inundație este mult mai probabil să se întâmple în anul $n + 1$ dacă s-a întâmplat în anul n. De asemenea, jucătorii de ruletă care fac greșeala să creadă că roșul este mult mai probabil să apară din nou dacă a apărut de mai multe ori la rând își formează credințele aplicând o *euristică a reprezentativității*. Aceștia cred – sau acționează ca și cum ar crede – că o succesiune scurtă de evenimente este probabil reprezentativă pentru o succesiune mult mai lungă, în care este inclusă.

De multe ori, oamenii nu înțeleg relația dintre procese întâmplătoare și distribuția rezultatelor. În timpul celui de-al Doilea Război Mondial, mulți londonezi erau siguri de faptul că germanii își concentrau bombardamentele în anumite zone ale orașului pentru că bombele cădeau în grupuri. Aceștia nu înțelegeau principiul statistic de bază conform căruia procesele aleatoare creează acumulări, deci că, dacă bombele ar fi căzut într-o manieră egal distribuită pe toată suprafața orașului, abia acest

lucru ar fi fost un semn că bombardamentele își aleg deliberat țintele. Un fapt care îi surprinde pe mulți dintre aceia care nu l-au mai întâlnit este că probabilitatea ca doi oameni dintr-un grup de doar 23 de persoane să aibă aceeași zi de naștere (zi și lună) este mai mare de 50%.

Gândirea magică

Să luăm acum în considerare diferite forme de *gândire magică*, adică tendința de a crede că putem exercita o influență cauzală asupra unor rezultate care se află în afara controlului nostru. De exemplu, oamenii vor paria mai mult pe o monedă care nu a fost încă aruncată decât pe una deja aruncată și pentru care rezultatul a fost anulat. În opera lui Proust, Robert Saint-Loup, prietenul naratorului, întreținea „un soi de credință superstițioasă că fidelitatea amantei sale față de el depinde de fidelitatea lui față de amantă". De asemenea, oamenii pot să nu înțeleagă distincția dintre *relevanța cauzală și cea de diagnoză*. Subiecții unui experiment care au fost făcuți să creadă că cel mai bun indicator al longevității este să își țină mâinile în apă rece ca gheața au stat cu mâinile în apă mai mult decât aceia cărora nu li s-a dat această (falsă) informație.[1] De asemenea, folosindu-și propriul comportament ca factor predictiv al comportamentului altora, oamenii pot alege strategia cooperării în Dilema Prizonierului ca și cum astfel ar putea să îi facă și pe ceilalți să coopereze. Într-un experiment, subiecții cooperanți care au fost rugați să prezică alegerea partenerului de interacțiune, precum și pe aceea a unui subiect cuplat cu altă persoană, aveau o probabilitate mult mai mare să atribuie cooperarea partenerului (și erau mult mai încrezători în predicția lor) decât subiectului cuplat cu altă persoană.[2]

Calvinismul pune la dispoziție un exemplu de gândire magică (capitolul 3). Dată fiind credința calvinistă în predestinare, s-ar părea că nu există niciun motiv pentru care un calvinist să nu se lase în voia plăcerilor lumești, care, conform presupoziției

---

[1] Această distincție dintre cauză și simptom nu este evidentă întotdeauna. Chiar și în 1959 marele statistician R.A. Fisher, presupunând existența unei trăsături genetice care îi predispune pe indivizi atât la fumat, cât și la cancerul de plămâni, a susținut că fumatul este un simptom pentru cancer, și nu cauza sa. (Ce-i drept, Fisher era plătit de anumite corporații producătoare de tutun la vremea respectivă.) Sau putem lua în considerare descoperirea discutată în capitolul 2 că, odată cu prelungirea perioadei de șomaj, scade și probabilitatea găsirii unui loc de muncă într-o perioadă determinată. Durata șomajului poate fi parte a diagnosticului capacității de a obține un loc de muncă sau ar putea să aibă o contribuție cauzală (prin demoralizare etc.) asupra șanselor de a găsi un loc de muncă.

[2] Discrepanța ne permite să excludem varianta că atribuirea unui comportament cooperant partenerului de interacțiune se poate datora doar „falsului efect al consensului" (capitolul 23).

inițiale, nu îi pot afecta soarta după moarte. Max Weber a susținut, însă, că învățătura calvinistă îi făcea pe adepții ei să adopte mai degrabă un stil de viață ascetic, nu pentru a dobândi mântuirea, ci pentru a dobândi certitudinea subiectivă de a fi printre cei aleși. Putem interpreta acest lucru în sensul că învățătura calvinistă face o confuzie între relevanța cauzală și cea de diagnoză a unui anumit tip de comportament. Acest lucru este evident într-o scrisoare pusă în circulație de baptiștii englezi în 1770: „Fiecare suflet care îl îmbrățișează pe Hristos pentru a fi salvat (...) trebuie încurajat. (...) Sufletul care vine spre Hristos nu trebuie să se teamă că nu va fi ales, pentru că numai cei aleși sunt gata să-L accepte." Dacă Dumnezeu a decis să fiu printre cei aleși, atunci a și *cauzat ca eu să vreau* să am un anumit tip de comportament.

Aceste erori (și multe altele, care au fost documentate pe larg) sunt în mare parte greșeli „reci" sau nemotivate, asemănătoare într-o oarecare măsură cu iluziile optice. Alte erori, cele „fierbinți", apar întrucât credințele agenților sunt motivate, adică sunt influențate într-un mod nedrept de dorințele acestora. După cum vom vedea în capitolul 11, influența cauzală a dorințelor asupra credințelor nu este intrinsec rațională. O dorință poate furniza un motiv pentru a investi o anumită cantitate de resurse în achiziționarea de informație. Informația astfel obținută poate servi ca motiv pentru a întreține o anumită credință. Cu toate că dorința nu furnizează un motiv pentru a întreține respectiva credință, aceasta intră într-un complex de formare a credințelor. Discrepanța dintre dorința inițială și credința finală este produsă de faptul că rezultatul căutării de informație este, prin definiție, necunoscut la momentul în care s-a luat decizia de a începe căutarea.

Formarea motivată de credințe

Influența dorințelor asupra credințelor, la care tocmai m-am referit, este, într-o manieră necontroversată, consistentă cu raționalitatea. O idee ceva mai controversată este aceea oferită de pariul lui Pascal. Așa cum am argumentat în capitolul anterior, Pascal susține că un agent care crede că există o probabilitate oricât de mică, dar diferită de zero, că Dumnezeu există ar trebui, din motivul pur instrumental de a-și maximiza valoarea așteptată, să încerce să dobândească credința fermă (la modul certitudinii) că Dumnezeu există întrucât, dacă acesta există, acea credință îi va asigura beatitudinea eternă. Premisele pentru acest argument sunt: 1) că credința certă asigură cu certitudine mântuirea și 2) că originea instrumentală a credinței nu îi influențează negativ eficacienta în vederea mântuirii. Cu toate că ambele premise pot fi dubioase din punct de vedere teologic, nu trebuie să ne preocupe acest subiect. Întrebarea este dacă această „decizie de a crede" este un proiect rațional. Într-un anumit sens, nu este: nu mă pot decide să cred oricând vreau, așa cum îmi pot ridica mâna oricând vreau. Dar se poate folosi o strategie indirectă de argumentare. Acționând *ca și cum* ai crede,

susține Pascal, vei sfârși prin a crede. Dar mecanismul prin care ar trebui să se întâmple așa ceva este destul de insondabil.[1]

Există și alte cazuri, în care cineva poate vrea să dobândească o credință pe care o consideră falsă pentru a beneficia de consecințele pe care le aduce întreținerea unei astfel de credințe. Dacă vreau să reduc consumul de alcool, dar îmi dau seama că nu sunt suficient de motivat de riscul de a deveni alcoolic, pot dori să cred că riscul este mai mare decât cred acum că este. Însă, în general, nu există niciun mecanism demn de încredere pentru a dobândi astfel de credințe. Dacă procesul în sine nu are o *componentă de autoeliminare*, prin care originarea credinței în dorința de a o dobândi să fie eliminată din gândirea conștientă, atunci dorința rămâne, cel mai probabil, o simplă aspirație.

În cazul „necontroversat", dorința agentului determină un anumit nivel de acumulare de informație, care va determina, la rândul său, o anumită credință. În cazul „controversat", dorința determină un comportament specific, care, la rândul său, va determina o credință specifică pe care agentul vrea să o întrețină. Ambele sunt strategii indirecte. Mă voi apleca acum asupra credințelor care sunt modelate de motivație în mod *direct*. Aceasta se poate întâmpla în două moduri, care corespund celor două trăsături de bază ale motivației: stimularea și conținutul. La fel cum spunem că piatra a spart gheața în virtutea greutății, și nu a culorii sale, tot astfel putem spune că o motivație afectează credința nu în virtutea conținutului său, ci în virtutea nivelului de stimulare care o acompaniază. Stimularea fiziologică moderată poate îmbunătăți calitatea formării de credințe prin concentrarea atenției și stimularea imaginației. „Când cineva știe că urmează să fie spânzurat într-o săptămână", spunea Dr. Johnson, „mintea sa devine foarte ascuțită." Totuși, funcția cognitivă se deteriorează dincolo de un anumit nivel de stimulare. În stări de foame extremă, de frică sau de poftă cauzată de dependență, e greu să gândești corect pentru că stimularea împiedică urmărirea pașilor unui raționament. Putem presupune că ascuțimea minții aceluia care urmează a fi spânzurat este destul de tocită cu o zi înainte de execuție. În testele academice de aptitudini o motivație foarte puternică de a ajunge la rezultatul corect poate cauza eroare, la fel cum dorința foarte puternică a unui pușcaș de a nimeri ținta îi poate face mâinile să tremure, astfel încât ratează.[2] În capitolul următor voi susține că, din cauza caracterului presant al emoțiilor, acestea îl pot determina pe agent să ocolească mecanismul normal, rațional al formării de credințe. Astfel, credințele pot fi *formate de motivație*, dar pot să nu fie *motivate*, întrucât agentul nu are nicio dorință deosebită de

---

[1] Din motive pe care le voi explica mai târziu, reducerea disonanței nu este un mecanism plauzibil.

[2] În astfel de cazuri, nu ajută să-ți spui să te calmezi sau să ți se spună acest lucru întrucât starea de a fi relaxat, la fel ca starea de somn, este esențialmente un produs derivat.

a crede că acestea sunt adevărate. Stimularea *începătoșează* mintea, dar nu *favorizează* o anumită credință în detrimentul alteia.

Raționalizarea

Credințele generate de conținut se împart în două clase. După cum am menționat mai devreme, agentul poate fi motivat să întrețină o *anumită credință* despre un subiect pentru că are nevoie de un răspuns sau din cauza intoleranței față de ignoranță. Agentul poate fi motivat să întrețină o anumită credință, cum ar fi aceea că partenera de viață este fidelă.[1] Cele mai importante mecanisme care generează această clasă de credințe sunt raționalizarea, gândirea deziderativă și autoînșelarea. Diferența dintre prima și ultimele două stă în relația cu comportamentul. În cazul raționalizării, comportamentul apare primul, fiind urmat de credință. (Ceea ce nu înseamnă că credințele, odată adoptate, nu pot determina mai departe un anumit comportament.) În cazul gândirii deziderative și al autoînșelării observăm o succesiune inversă.

Ca un exemplu de raționalizare, putem lua în considerare un experiment-standard care pune în lumină „disonanța cognitivă". Subiecților din două grupuri li se cere să scrie un eseu prin care să ofere argumente referitoare la problema avortului și care să vină în sprijinul poziției pe care *nu* o favorizează. Subiecții din primul grup sunt plătiți cu o sumă considerabilă de bani pentru participare, în timp ce subiecții din al doilea grup sunt rugați să participe doar de dragul de a face un serviciu aceluia care implementează experimentul. După ce redactează eseul, cei din al doilea grup, dar nu și cei din primul, manifestă o atitudine mai favorabilă poziției pentru care au argumentat. Explicația plauzibilă este că toți subiecții doresc să aibă un *motiv* pentru ceea ce fac. Membrii primului grup pot invoca doar motivul banilor.[2] Membrii celui de-al doilea grup pot menționa credințele lor (ajustate) ca motiv pentru punctul de vedere în favoarea căruia argumentează.[3]

Conform unui proverb franțuzesc, „Cine a ofensat nu poate ierta". Dacă am făcut rău în mod nedrept unei alte persoane, s-ar putea să nu fiu capabil să admit față de mine însumi că am făcut ceva rău. În loc să fac acest lucru, voi căuta un defect la cealaltă persoană sau măcar scuză pentru comportamentul meu. Violatorii se justifică:

---

[1] Sau, ca în cazul lui Othello, că nu este fidelă. Astfel de „credințe contramotivaționale" au fost menționate pe scurt în capitolul 2 și vor fi discutate în capitolul 23.

[2] Pedeapsa pentru neparticipare ar fi și ea un motiv suficient. Aceasta explică de ce, în timpul regimului comunist, cetățenii puteau duce o viață dublă, fără ca lucrurile pe care le respingeau în forul lor interior să fie subminate de entuziasmul pe care îl jucau în viața publică.

[3] În pariul lui Pascal, motivul pentru a acționa ca și cum ai crede este atât de copleșitor – perspectiva beatitudinii eterne –, încât credinciosul nu are nevoie să caute altă explicație pentru comportamentul său.

„Era îmbrăcată prea provocator", scuză la care uneori pot subscrie și judecătorii. Cei care se angajează în violențe antisemite vor inventa o poveste conform căreia, din moment ce evreii își datorează succesul unor mijloace imorale sau ilegale, atunci merită să fie pedepsiți. Comportamentul determină credințe, și nu invers, lucru care se poate deduce din flexibilitatea extraordinară a raționalizării. După cum am observat în capitolul 3, retorica antisemită include caracterizarea evreilor drept „paraziți", precum și teza omnipotenței evreilor în istorie.

Gândirea dezideративă

Mă voi ocupa în cele ce urmează de gândirea dezideративă și de autoînșelare. Aceste două fenomene, puțin înțelese, au în comun faptul că dorința ca *p* să se adevereasă cauzează credința că *p* s-a adeverit. În cazul gândirii dezideративe, acest proces are un singur pas: dorința este mama gândului. Probele nu sunt negate, ci pur și simplu ignorate. Prin urmare, credința formată prin dorință ar putea fi una și aceeași cu credința justificată prin probe, dacă ar fi fost consultate.[1] Autoînșelarea, așa cum este concepută de obicei, implică patru pași: primul presupune consultarea probelor, al doilea presupune formarea unei credințe potrivite, al treilea presupune respingerea credinței sau suprimarea ei pentru că nu este consistentă cu dorința noastră, iar în cadrul ultimului pas, dorința cauzează formarea unei alte credințe, mult mai acceptabile, care urmează să îi ia locul. Autoînșelarea este un fenomen paradoxal, a cărui existență și chiar posibilitate a fost pusă la îndoială, așa că voi începe cu problema mai simplă, a gândirii dezideративe.

---

[1] Ignorarea acestui aspect poate fi sursa formării iraționale de credințe. Întrucât este ușor să detectăm în alții modul în care operează formarea de credințe motivate, avem tendința de a fi neîncrezători în concluziile la care ajungem astfel, fără să confruntăm probele care le-ar putea veni în sprijin. Până prin 1990, credeam, alături de cei mai mulți dintre prietenii mei, că, pe o scală a răului de la 0 la 10 (10 însemnând cel mai malefic), comunismul se află undeva la un 7 sau 8. În lumina ultimelor descoperiri, tind să cred că 10 este mai potrivit. Motivul pentru care am interpretat greșit probele nu este credința idealistă că regimul comunist reprezintă un ideal care merită urmărit, dar care a fost compromis de comuniștii înșiși. Într-un asemenea caz, aș fi fost o victimă a gândirii dezideративe sau a autoînșelării. Mai degrabă, pot spune că am fost indus în eroare de temperamentul isteric al tuturor acelora care susțineau răutatea extremă a comunismului. Ignoranța mea în raport cu afirmațiile acestora nu era în întregime irațională. În medie, este justificat să nu ții cont de afirmațiile care ți se livrează într-un mod isteric. Însă chiar și istericii pot avea dreptate, chiar dacă din motive greșite. Pentru că atunci simțeam, și cred și acum, că toți acești anticomuniști isterici ar fi susținut cauza lor cu aceeași tărie, indiferent de probele care le-ar fi fost puse la dispoziție, nu credeam că afirmațiile lor corespundeau în totalitate probelor. Am făcut greșeala de a-i considera asemenea unui ceas care este tot timpul înainte cu o oră, și nu asemenea ceasului stricat care indică ora exactă de două ori pe zi.

Înainte de a sugera un mecanism responsabil pentru gândirea dezideravă, vreau să menționez faptul că existența sa nu poate fi negată, spre deosebire de autoînșelare. Se poate nega că apare în situații cu miză mare sau că afectează comportamente agregate, precum bursele de valori sau alegerile politice, dar nu i se poate nega existența. Mai mult, numeroase credințe formate în mod iluzoriu servesc ca premise pentru acțiune, fiind astfel mai mult decât niște „cvasi-credințe". Fumătorii care se amăgesc în a crede că fumatul nu este dăunător sănătății, în general sau doar în cazul lor particular, s-ar fi lăsat de fumat sau ar fi încercat dacă ar fi nutrit credințe mai raționale.[1] Indivizii foarte încrezători în sine, care cred în mod iluzoriu că sunt mult mai capabili decât sunt de fapt, se pot aventura în întreprinderi pe care le-ar fi evitat altminteri. Oamenii care se amăgesc gândindu-se că au același succes ca alții își pot pierde imboldul de a-și îmbunătăți abilitățile. Mecanismul obișnuit este următorul: în primul rând, persoana este motivată să creadă că are succes. În al doilea rând, aceasta găsește anumite dimensiuni ale vieții sale în care nu o duce chiar atât de rău. În al treilea rând, sporește importanța acelor zone pentru a-și putea spune că are succes la nivel global. În cel de-al patrulea rând, își relaxează eforturile de a obține succes și în alte aspecte ale vieții.

Pentru a naviga în viață este util, din punct de vedere instrumental, să ai credințe corecte. În același timp, credințele pot fi în mod intrinsec plăcute sau neplăcute, adică pot cauza emoții pozitive sau negative. Dacă mi se spune că am cancer, pot căuta un tratament, dar credința în sine mă va face să mă simt groaznic. După cum spune Freud, aceia care sunt guvernați de principiul realității caută credințele corecte, pe când aceia care sunt supuși principiului plăcerii caută credințele plăcute. Această distincție se aplică doar credințelor propriu-zise, nu și cvasi-credințelor. Oamenii care își formează credințe nerealiste, cum ar fi aceea că vor primi un premiu mare în bani pentru reușitele lor, dar care, totuși, nu cheltuiesc banii înainte să îi primească, sunt, în cel mai rău caz, supuși unei forme inofensive a principiului plăcerii. În varianta nocivă a principiului, convingerea că vor primi premiul îi determină să se îndatoreze.

Formarea de credințe poate avea și costuri. Dacă credințele apar din cauza beneficiilor lor intrinseci, costul este acela că renunț la beneficiile instrumentale. Acest cost depinde atât de rezultatul care va apărea dacă credința motivată este falsă, prin comparație cu ce s-ar întâmpla dacă se adeverește, cât și de probabilitatea să fie falsă. Voi numi aceste componente, *componenta-rezultat* și *componenta-probabilitate* a costurilor. Pe de altă parte, credințele care sunt formate pentru a furniza beneficii instrumentale pot avea costul că renunț la cele intrinseci. Voi discuta unele cazuri în care formarea rațională de credințe are costuri intrinseci pe termen scurt și beneficii instrumentale pe termen lung, în timp ce credința motivată are beneficii intrinseci pe termen scurt

[1] Ca și în cazul alcoolului, renunțarea poate necesita credința irațională că fumatul este *mult mai* periculos decât este de fapt.

și costuri instrumentale pe termen lung. Prin aceasta mă opun explicației formării credințelor motivate propuse de economiști și psihologi, argumentând că trebuie să ne bazăm pe amândouă.

Economiștii pun accent pe *costuri*. Unii dintre ei au susținut, de pildă, că muncitorii își formează credințe motivate despre siguranța la locul de muncă dacă beneficiul de a întreține credința depășește costurile. Dacă beneficiul psihologic al suprimării fricii provocate de o anumită activitate depășește costul dat de șansele sporite de accident, atunci muncitorul va crede că activitatea pe care o desfășoară este sigură. De aici rezultă că agentul sondează inconștient probele pentru a vedea dacă își permite să adopte credința motivată. Mai departe, se presupune că nu există constrângeri care să împiedice formarea credinței: muncitorul poate crede orice alege, indiferent de informațiile disponibile. Cred că acesta este un model greșit, nu doar pentru că nu admite constrângeri, dar și pentru că se aduc în discuție costurile într-un mod greșit. Beneficiile credințelor motivate se manifestă acum, iar costurile posibile doar mai târziu. Pentru ca argumentul să fie bun, trebuie să presupunem că gândirea inconștientă este capabilă să realizeze astfel de compensații intertemporale. După cum am argumentat în introducerea la Partea a II-a, nu există probe care să susțină o astfel de idee.

O idee mai plauzibilă, în opinia mea, este aceea că gândirea dezideративă este declanșată de *componenta-rezultat cunoscută a costurilor*. Când miza este mică, agentul își poate forma o credință fără să se gândească la componenta probabilitate a costurilor, adică fără să ia în considerare probele. Nu există nicio compensație în acest caz, ci doar un proces compus din doi pași. Primul presupune că agentul ia în considerare toate mizele. Dacă acestea sunt mici, adoptă credința mai lăcută. Dacă acesta sunt mari, ia în considerare probele și, dacă este necesar, acumulează cât mai multe. Ideea că mizele sunt „mici" variază, firește, de la individ la individ. Tot ce se poate spune este că, pentru un anumit individ și dacă totul merge cum trebuie, gândirea dezideративă este mult mai probabilă atunci când mizele sunt mici.

Psihologii se concentrează pe *constrângeri*. Un agent care se apucă de fumat poate fi tentat să își formeze o credință dezideративă și să creadă că fumatul nu este periculos sau, cel puțin, nu este periculos pentru el. Procedând astfel, agentul poate fi constrâns de credințele formate în prealabil privitoare la pericolele aferente fumatului. Prima oară când o persoană are parte de un eșec la un examen, se poate minți că a avut ghinion, dar, dacă rezultatul se repetă și la următoarele patru susțineri, povestea este puțin probabil să mai funcționeze dacă pică și a șasea oară. Putem să revenim la exemplul cu spectacolele scumpe de pe Broadway, pe care l-am introdus în capitolul 1. Dacă am plătit 75 de dolari pentru un bilet, iar spectacolul se dovedește oribil, amintirea sumei pe care am plătit-o va fi, probabil, prea vie în mintea mea pentru a fi supusă unei revizuiri dezideративe în sens descendent. Dar, dată fiind natura intangibilă și multidimensională a aprecierilor estetice, este mult mai ușor să ajustez ascendent evaluarea

spectacolului. Similar, cu toate că există probe că evenimentele mai probabile sunt percepute ca fiind mai dezirabile și că evenimentele dezirabile sunt percepute ca fiind mai probabile, ultimul efect este mult mai constrâns decât primul.

Într-un experiment instructiv, subiecții se așteptau să participe la un joc de cultură generală pe teme de istorie alături de o anumită persoană fie ca partener, fie ca adversar. După expunerea unei mostre a abilităților persoanei, în care aceasta răspunde corect la toate întrebările, cei care se așteptau ca persoana să le fie partener (și își doreau, prin urmare, ca această să aibă abilități remarcabile) au evaluat persoana ca fiind mai bună la istorie decât aceia care se așteptau ca persoana să le fie adversar (și care își doreau ca aceasta să aibă abilități limitate). În același timp, subiecții erau constrânși de natura informației pe care o primeau, din moment ce chiar subiecții care se așteptau ca persoana să le fie adversar au evaluat-o ca fiind peste medie. Experimentul este limitat prin aceea că nu le-a oferit subiecților șansa de a *acționa* în baza acestor credințe, cu consecința costisitoare care ar putea decurge din subestimarea unui adversar. Din câte știm, putem avea de-a face cu simple cvasi-credințe.

În exemplele precedente, gândirea dezideративă este constrânsă de credințe anterioare cu privire la fapte. În alte cazuri, gândirea dezideративă poate fi constrânsă de credințe cauzale plauzibile. Gândirea dezideративă presupune, de multe ori, „să-ți spui o poveste", ideea unei povești fiind foarte apropiată de ideea unui mecanism, pe care am discutat-o în capitolul 2. Abundența de mecanisme facilitează găsirea unei istorisiri anume care poate justifica orice credință pe care o vrem adevărată. Pot respinge un zvon neconvenabil cu ajutorul proverbului „Zvonurile sunt îndeobște mincinoase" și pot accepta un zvon convenabil cu ajutorul proverbului „Zvonurile sunt rareori mincinoase". Să presupunem că citesc în ghidul de admitere la o școală de asistență socială că stabilitatea emoțională constituie o condiție pentru oamenii care caută o astfel de profesie. Dacă mama mea și-a părăsit locul de muncă pentru a avea grijă de mine după ce m-am născut, pot să-mi întăresc credința în stabilitate spunându-mi o poveste conform căreia copiii trag numai foloase de pe urma atenției permanente a părinților lor. Dacă și-a păstrat locul de muncă și m-a trimis la creșă, pot adopta o istorisire conform căreia copiii beneficiază de pe urma socializării cu alți copii și a unor părinți care cunosc împlinirea profesională în afara căminului.[1] Dacă echipa mea favorită de fotbal joacă prost, îmi pot menține credința în superioritatea acesteia față de adversar prin (ceva ce poate fi conceput ca) un eveniment norocos. „Dacă mingea n-ar fi fost deviată de arbitru, atacantul de pe extremă ar fi primit pasa și ar fi putut șuta pentru gol." În cazul în care calul pe care am pariat termină pe locul doi, îmi pot menține credința în abilitățile mele de parior spunând că „era cât pe ce să câștig". Într-o mostră

[1] De fapt, nu s-au descoperit diferențe consistente în dezvoltarea ulterioară a copiilor crescuți în aceste două medii.

de gândire dezideravă flagrant irațională, dacă pariez pe 32 și rezultatul este 33, pot spune că „era cât pe ce să câștig", chiar dacă numerele acestea sunt departe unul de altul pe ruletă.[1]

Totuși, uneori, nu avem istorisiri la îndemână. Să presupunem că o persoană pune bani pe 24. Rezultatul este 15, număr apropiat de 24 pe ruletă; prin urmare, i se confirmă credința în abilitățile de jucător. Probabil că și alte rezultate ar fi fost luate drept confirmări, întrucât și 5, 10 sau 33 sunt aproape de 24. De asemenea, ar putea considera drept confirmări și numerele 22, 23, 25 și 26 pentru că valoarea lor numerică este apropiată sau ar putea să considere numerele 20, 21, 26 și 27 pentru că sunt adiacente pe masă. Astfel, 13 din 37 de rezultate posibile ar putea fi luate drept confirmări ale talentului de a paria. Însă, rezultă de aici că rămân 24 de rezultate posibile pentru care nu este disponibilă nicio istorisire. În cazul unui astfel de rezultat, chiar și o persoană predispusă la gândire dezideravă și foarte motivată să adopte o credință specifică ajunge să se confrunte cu realitatea.

Autoînșelarea

Să luăm acum în considerare chestiunea autoînșelării. În viața de zi cu zi, oamenii par a se amăgi cu privire la lucruri precum greutatea, sănătatea, cât de mult beau, tendința de a lăsa lucrurile pe mâine sau fidelitatea partenerilor de viață. Într-un scenariu destul de des întâlnit, aceștia primesc o informație cum că ceva nu este în ordine, iar apoi nu reușesc să facă nimic pentru a ajunge la o concluzie clară. Uitându-mă în oglindă, îmi dau seama că am luat în greutate, dar nu îmi pot da seama cât de mult. Abținându-mă să mă urc pe cântar, îmi pot spune că este vorba doar despre câteva kilograme, pe care le pot da jos dacă vreau. O femeie simte un nodul într-un sân, dar nu reușește să se programeze pentru a verifica dacă este benign sau malign. În astfel de cazuri, autoînșelarea este favorizată de lipsa unor cunoștințe exacte. Femeia nu trage mai întâi concluzia că are cancer pentru ca apoi să suprime această credință. Mai degrabă, ea suspectează că *ar putea* avea cancer.

Acest caz, pe care îl voi considera drept paradigmatic, are următoarele caracteristici:

1. Suspiciunea inițială de cancer ia forma unei credințe cu o probabilitate mică.
2. Aceasta este însoțită de credința fermă că, deși are cancer și nu face nimic în această privință, rezultatul este, mai mult ca sigur, fatal.

---

[1] În același timp, oamenii pot fi mai dezamăgiți dacă numărul lor a fost foarte aproape de numărul câștigător. Unele loterii naționale oferă mici „premii de consolare" pentru aceia care „au fost cât pe ce să câștige".

3. Este, de asemenea, însoțită de credința fermă că, deși are cancer și face ceva în această privință, rezultatul poate totuși fi fatal, iar, dacă nu este, tratamentul va fi foarte neplăcut.
4. Femeia nu stă să se întrebe în niciun moment dacă durerea provocată de tratament (comparată cu durerea provocată de boala netratată) nu compensează riscul diferențial al mortalității.
5. În loc de aceasta, femeia refuză vizita la medic pentru a stabili dacă are sau nu cancer.

În acest exemplu, trăsăturile cruciale sunt (1) și (3). Întrucât credința inițială are o probabilitate foarte redusă, costurile de ajustare sunt foarte mici. Femeia își poate îndrepta atenția cu ușurință asupra unui număr mare de istorisiri pe care le va fi auzit despre gâlme inofensive și despre panici nejustificate. *Și totuși, în lipsa trăsăturii (3), aceasta nu are vreun motiv pentru a-și remodela credințele.* Dacă ar ști că există un tratament care nu implică nici costuri, nici durere, dar care i-ar aduce vindecarea, femeia nu ar avea niciun motiv să nu meargă la medic. Nu se cunosc forme de iraționalitate care să favorizeze tendința de a bloca evitarea *care nu implică vreun cost* a dezastrelor cu probabilitate redusă prin, de pildă, transformarea unei probabilități foarte mici într-o probabilitate zero.

Astfel concepută, autoînșelarea nu implică întreținerea simultană a două credințe contradictorii, una conștientă, iar alta inconștientă. Când evaluarea inițială de probabilitate este înlocuită cu alta, cea dintâi nu este trimisă în inconștient, ci *dispare definitiv*. Mulți autori consideră că această contradicție este trăsătura cea mai importantă a autoînșelării. Însă lucrurile nu trebuie să stea neapărat astfel. Persoana care se autoînșală nu este (neapărat) ca persoana care detestă să vadă pisici, dar își dă seama că, pentru a-și întoarce privirea de la ele, trebuie mai întâi să le observe. Este, mai degrabă, precum persoana care vede o umbră în întuneric care ar putea fi o pisică, dar care ar putea la fel de bine să fie orice altceva. Aversiunea sa față de pisici poate fi satisfăcută reinterpretând mai întâi umbra, iar apoi rămânând pe loc pentru a verifica dacă este chiar o pisică.

Mă îndoiesc că psihologia sau filosofia au șanse mai mari să ne spună mai multe despre autoînșelare decât literatura. În *Swann*, Proust descrie modul în care Swann se gândește „la vremurile când lumea i-o descria pe Odette [amanta sa] drept o femeie întreținută" și cât de deplasată i se părea o astfel de descriere în raport cu Odette așa cum o cunoștea el. Printr-o serie de asocieri mentale, acesta ajunge să se gândească la bancherul său, amintindu-și că trebuie să scoată niște bani din bancă pentru a o ajuta pe femeie, aflată în anumite dificultăți materiale.

Apoi, dintr-odată, se întreabă dacă nu cumva chiar asta se înțelege prin „a o întreține"(...) și dacă nu cumva i se potrivesc Odettei (...) acele cuvinte pe care le găsea atât de nepotrivite ei – „femeie întreținută". Însă Swann nu urmează ideea în profunzime pentru că un atac de lene a minții, care, în cazul său, era congenitală, intermitentă și providențială, îi stinse în chiar acel moment toată lumina inteligenței, la fel de abrupt precum, mai târziu, când lumina electrică avea să fie instalată peste tot, puteai întrerupe cu un simplu gest curentul în toată casa. Cu mintea bâjbâind o vreme prin întuneric, își dădu ochelarii jos, își șterse lentilele, își trecu mâna prin fața ochilor și revăzu lumina doar când se pomeni în prezența unei idei complet diferite, anume că ar trebui să încerce să-i trimită Odettei șase sau șapte sute de franci luna viitoare, în loc de cinci sute, întrucât asta ar fi o surpriză care i-ar face mare plăcere.

◈◈◈

Notă bibliografică

Multe dintre descoperirile la care se face referire în acest capitol pot fi documentate în următoarele surse: D. Kahneman, P. Slovic și A. Tversky (editori), *Judgment Under Uncertainty* (Cambridge University Press, 1982), D. Bell, H. Raiffa și A. Tversky (editori), *Decision Making* (Cambridge University Press, 1988), T. Connolly, H. Arkes și K.R. Hammond (editori), *Judgment and Decision Making* (Cambridge University Press, 2000), D. Kahneman și A. Tversky (editori), *Choices, Values and Frames* (Cambridge University Press, 2000), T. Gilovich, D. Griffin, și D. Kahneman (editori), *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, 2002), C. Camerer, G. Loewenstein și M. Rabin (editori), *Advances in Behavioral Economics* (New York, Russell Sage, 2004), I. Brocas și J. Carillo (editori), *The Psychology of Economic Decision*, vol. 1 și 2 (Oxford University Press, 2003, 2004). Pentru documentarea tendinței de a atribui cooperarea mai degrabă partenerilor decât altor persoane se poate vedea L. Messé și J. Sivacek, „Predictions of others' responses in a mixed-motive game: Self-justification or false consensus?", *Journal of Personality and Social Psychology*, 37 (1979), pp. 602-607. Dubla incompetență a ignorantului este documentată în J. Kruger și D. Dunning, „Unskilled and unaware of it", *Journal of Personality and Social Psychology*, 77 (1999), pp. 1121-1134. Un studiu sofisticat asupra judecăților lipsite de credibilitate ale unor experți („aricii") și asupra calității, oarecum mai credibile, a judecăților altora („vulpile") se află în P. Tetlock, *Expert Political Judgment* (Princeton, NJ, Princeton University Press, 2005). Pentru date despre cutremure și inundații, vezi P. Slovic, *The Perception of Risk* (Sterling, VA, Earthscan, 2000). Pentru logica (ilogică) a teoriilor conspiraționiste,

vezi, B. Keeley, „Of conspiracy theories", *Journal of Philosophy*, 96 (1999), pp. 109-126. Pentru teoriile foametei vezi S. Kaplan, „The famine plot persuasion in eighteenth-century France", *Transactions of the American Philosophical Society*, 72 (1982), și F. Ploux, *De bouche à oreille: Naissance et propagation des rumeurs dans la France du XIXe siècle* (Paris, Aubier, 2003). Un studiu privitor la gândirea conspiraționistă se află în R. Hofstadter, *The Paranoid Style in American Politics* (Cambridge, MA, Harvard University Press, 1964). Rolul acesteia în Orientul Mijlociu este subiectul cărții lui D. Pipes, *The Hidden Hand: Middle East Fears of Conspiracy* (New York, St. Martin's Press, 1998). Comentariile mele despre modul în care doi economiști (George Akerlof și Matthew Rabin) și un psiholog (Ziva Kunda) abordează formarea credințelor motivate sunt prezentate pe larg în „Costs and constraints in the economy of the mind", în I. Brocas și J. Carillo (editori), *The Psychology of Economic Decisions*, vol. 2 (Oxford University Press, 2004). Cea mai bună trecere în revistă a autoînșelării se găsește într-o ediție specială din *Behavioral and Brain Sciences*, 20 (1997), organizată în jurul unui articol de A. Mele „Real self-deception".

BCU IASI CENTRAL UNIVERSITY LIBRARY

Capitolul 8

EMOȚIILE

Rolul emoțiilor

Emoțiile se manifestă în viață în trei moduri. În punctul culminant a intensității, acestea sunt cele mai importante *surse ale fericirii și nefericirii*, umbrind cu desăvârșire plăcerile hedoniste și durerea fizică. Dragostea radioasă a Annei Elliott la sfârșitul cărții *Persuasiune* constituie fericirea absolută. Pe de altă parte, emoția rușinii poate fi devastatoare. Voltaire scrie: „A fi disprețuit de cei de alături este un lucru pe care nimeni nu a fost și nu va fi vreodată în stare să îl îndure."

Rușinea ilustrează și al doilea mod în care se manifestă emoțiile, și anume prin *impactul lor asupra comportamentului*. În capitolul 4 am menționat o serie de cazuri în care oamenii se sinucid din cauza emoției copleșitoare a rușinii. În acest capitol mă voi ocupa mai ales de *tendințele spre acțiune* care sunt asociate emoțiilor. Măsura în care aceste tendințe sunt transformate în comportamente manifeste va forma obiectul preocupărilor noastre în capitole ulterioare.

În al treilea rând, emoțiile sunt importante din cauza impactului lor asupra *altor stări mentale*, în special asupra credințelor. Când dorința ca o anumită stare să se realizeze este susținută de o emoție puternică, tendința de a crede că starea se și realizează poate fi irezistibilă. Iată ce spune Stendhal în *Despre iubire*: „Din momentul în care se îndrăgostește, nici măcar cel mai înțelept dintre oameni nu mai vede lucrurile *așa cum sunt în realitate*. (...) Acesta nu mai îngăduie existența întâmplării în lucruri și își pierde simțul probabilului; judecând doar după modul în care îi este afectată fericirea, orice își imaginează acesta devine realitate." În *A la recherche du temps perdu*, Proust urmărește aceeași temă pe sute de pagini, cu mai multe variațiuni și întorsături de situație decât am fi crezut posibil.

Ce sunt emoțiile?

Înainte să considerăm mai în detaliu fiecare aspect al emoției, trebuie să fac câteva precizări despre *ce anume sunt emoțiile* și despre *ce fel de emoții există*. Nu avem

o definiție unanim acceptată a ceea ce înțelegem prin emoție, altfel spus, nu s-a convenit asupra unei liste de condiții necesare și suficiente care să caracterizeze noțiunea de emoție. Nu există nici măcar o listă acceptată de condiții necesare. Cu toate că voi pune în discuție un număr considerabil de trăsături comune ale stărilor pe care le înțelegem în mod preanalitic ca fiind emoții, se pot da contraexemple pentru fiecare dintre acestea. Altfel spus, pentru orice astfel de trăsătură există anumite emoții sau manifestări ale unor emoții în care aceasta lipsește. Ne putem gândi că tendințele spre acțiune sunt cruciale pentru emoții, dar emoțiile estetice ne pun în fața unui contraexemplu. Ne putem imagina că „un timp de înjumătățire foarte scurt", adică tendința de a dispărea rapid, este o trăsătură esențială a emoției, dar, în anumite cazuri, dragostea romantică neîmpărtășită (ca în cazul lui Cyrano de Bergerac) sau dorința pătimașă de răzbunare se poate întinde pe ani întregi. Ne putem gândi că emoțiile sunt declanșate de credințe, dar atunci cum putem explica faptul că oamenii se pot emoționa citind diverse istorisiri sau urmărind filme care sunt, evident, fictive? Putem da multe alte exemple de trăsături despre care ne-am putea gândi că sunt esențiale, dar care se dovedesc a lipsi în anumite cazuri.[1]

Ținând cont de această problemă, un răspuns firesc ar fi să negăm că „emoția" este o categorie utilă din punct de vedere științific. În limbajul filosofilor, emoțiile nu par a forma un *gen natural*. În ciuda diferențelor, balenele și liliecii, ca mamifere, fac parte din același gen natural. Balenele și rechinii, în ciuda asemănărilor, nu fac parte din același gen, și nici liliecii și păsările. Mânia și iubirea au în comun capacitatea de a încețoșa mintea și de a o face părtinitoare, dar această asemănare nu le transformă într-un gen natural. Pentru a vedea modul în care raționamentul prin analogie ne poate duce într-o direcție greșită, putem observa că administrarea de amfetamine și dragostea romantică au multe efecte comune: conștiență acută, un nivel ridicat de energie, scăderea nevoii de somn și mâncare, precum și sentimente de euforie. Cu toate acestea, presupun că nimeni nu va susține că aceste două stări formează un gen natural.[2]

Pentru scopul explicației științifice, această enigmă poate rămâne fără soluție. Ne putem concentra atenția asupra manifestărilor emoționale în care un anumit număr de trăsături sunt observate cu regularitate și ne putem întreba cum ne pot ajuta acestea pentru a explica stări mentale sau comportamente. Faptul că, în alte manifestări pe care le-am considera în mod intuitiv emoții, unele dintre aceste trăsături lipsesc este o chestiune interesantă din punct de vedere conceptual, dar care nu diminuează

---

[1] În scopuri care țin de diagnoză, se poate stipula că o stare mentală este o emoție dacă are (să zicem) opt sau mai multe dintre douăsprezece caracteristici. În scopuri explicative, avem nevoie, însă, de mai mult de atât.

[2] Însă este posibil să acapareze o mare parte din același circuit neuronal.

eficiența lor explicativă acolo unde aceste trăsături sunt prezente. Caracteristicile asupra cărora vreau să atrag atenția sunt următoarele:

- *Antecedente cognitive.* Emoțiile sunt declanșate de credințe, de cele mai multe ori de faptul că agentul dobândește o nouă credință. Emoțiile pot avea și alte condiții cauzale (suntem mult mai iritabili dacă suntem obosiți), dar prezența acestora nu este suficientă pentru ca emoția să apară, la fel cum un drum alunecos nu va produce de unul singur un accident de mașină.
- *Stimulare fiziologică.* Emoțiile merg mână în mână cu modificări ale pulsului, ale conductivității electrice la nivelul pielii, ale temperaturii corpului, ale tensiunii arteriale, ale respirației și ale altor variabile.
- *Expresii fiziologice.* Emoțiile sunt însoțite de semne observabile caracteristice, precum postura corpului, tonul vocii, culoarea obrajilor (înroșirea provocată de jenă), zâmbetul sau dezvelirea dinților, râsul și chicotitul, lamentarea și plânsul, precum și mânia albă sau roșie (manifestată prin paloare sau, respectiv, prin înroșirea feței).
- *Tendințe spre acțiune.* Emoțiile sunt însoțite de tendințe sau nevoi de a performa anumite acțiuni. Cu toate că aceste tendințe nu trebuie neapărat să ducă la un comportament manifest, acestea sunt mult mai mult decât simple dispoziții – sunt forme de comportament incipient, și nu doar forme potențiale de comportament.
- *Obiecte intenționale.* Spre deosebire de alte fenomene viscerale, cum ar fi durerea sau foamea, emoțiile sunt *despre* ceva. Acestea pot avea „obiecte propoziționale" („Sunt indignat că...") sau obiecte non-propoziționale („Sunt indignat cu privire la...").
- *Valența.* Acesta este un termen tehnic pentru dimensiunea durere-plăcere a emoțiilor și a modului în care le experimentăm. După cum s-a observat, valența poate merge de la fericirea radioasă a Annei Elliott până la rușinea zdrobitoare a demascării publice a consumatorilor de materiale pornografice cu minori.

Însă nu produc oare emoțiile, la fel ca și culorile, *sentimente* calitative specifice? Rușinea și vina, de exemplu, par a fi simțite într-un mod foarte distinct, pe care nu îl putem reduce la faptul că rușinea este neplăcută într-un grad mai mare. S-a pus în evidență faptul că este suficient să inserezi un electrod în creierul uman pentru ca subiectul să se simtă trist, rușinat sau temător, chiar dacă nu poate identifica o cauză sau un obiect al sentimentului. Oricât de important se va dovedi acest aspect cu privire la

înțelegerea emoției, nu este încă suficient de bine înțeles pentru a sugera ipoteze cauzale specifice.

Ce fel de emoții există?

Voi enumera și voi descrie pe scurt vreo douăzeci și ceva de emoții, fără să susțin că această clasificare este mai bună decât acelea care au fost propuse până acum. Scopul meu este să arunc o anumită lumină asupra emoțiilor care au o importanță intrinsecă sau cauzală în viața socială, și nu de a satisface preocupările (legitime) ale teoreticienilor emoției. Nu am nimic de spus despre clasificarea emoțiilor în „fundamentale" și „derivate".

Un grup important de emoții sunt emoțiile evaluative. Acestea presupun o evaluare pozitivă sau negativă a caracterului sau comportamentului propriu sau al altcuiva.[1] Dacă o emoție este declanșată de comportamentul altei persoane, acel comportament poate fi direcționat fie spre sine, fie spre un terț. Aceste distincții ne pun în fața a zece (sau unsprezece) emoții:

- *Rușinea* este declanșată de o credință negativă despre caracterul propriu.
- *Disprețul* și *ura* sunt declanșate de sentimente negative despre caracterul altcuiva. Disprețul este provocat de gândul că celălalt este inferior, iar ura de faptul că celălalt este rău.
- *Vina* este declanșată de o credință negativă despre acțiunea proprie.
- *Mânia* este declanșată de o credință negativă despre acțiunea altcuiva asupra propriei persoane.
- *Indignarea carteziană*[2] este declanșată de o credință negativă despre acțiunea altcuiva asupra unui terț.
- *Aroganța* este declanșată de o credință pozitivă despre caracterul propriu.
- *Aprecierea* este declanșată de o atitudine pozitivă despre caracterul altcuiva.
- *Mândria* este declanșată de o credință pozitivă privitoare la propria acțiune.
- *Recunoștința* este declanșată de o credință pozitivă despre acțiunea altcuiva asupra propriei persoane.
- *Admirația* este declanșată de o credință pozitivă despre acțiunea altcuiva asupra unui terț.

---

[1] Emoțiile declanșate de evaluări negative ale propriei persoane au întotdeauna o valență negativă. Cele cauzate de evaluări negative ale altora sunt mai ambigue din acest punct de vedere.

[2] Emoția a fost identificată, pentru prima dată, de Descartes, care a făcut și precizarea foarte importantă că, atunci când agentul simte *iubire* față de un terț, reacția este una de mânie, mai degrabă decât una de indignare.

În al doilea rând, emoțiile sunt generate de gândul că altcineva este în posesia meritată sau nemeritată a unui lucru bun sau rău.[1] Ținta acestor emoții nu este nici acțiunea individuală, nici caracterul individului, ci o stare de lucruri. Urmând cele spuse de Aristotel în *Retorica*, putem distinge șase (sau șapte) cazuri.

- *Invidia* este cauzată de bunul meritat al altcuiva.
- *Indignarea aristotelică* este cauzată de bunul nemeritat al altcuiva.[2] Această emoție, înrudită îndeaproape cu *resentimentul*, este cauzată de inversarea unei ierarhii a meritului, când un grup sau un individ până atunci inferior apare ca dominant.
- *Simpatia* este cauzată de bunul meritat al altcuiva.
- *Mila* este cauzată de răul nemeritat al altcuiva.
- *Răutatea* este cauzată de răul nemeritat al altcuiva.
- *Bucuria de răul altuia* este cauzată de răul meritat al altcuiva.

În al treilea rând, există emoții pozitive sau negative, generate de gândul la lucrurile bune sau rele care s-au întâmplat sau se vor întâmpla propriei persoane – bucuria și mâhnirea, cu varietățile și categoriile lor înrudite. După cum au observat mulți gânditori, evenimentele negative din trecut pot genera, de asemenea, emoții pozitive în prezent, iar evenimentele bune pot genera emoții negative. Astfel, în antologia principală de maxime antice, *Sentențele* lui Publilius Syrus, se poate găsi atât „Reamintirea pericolelor trecute este plăcută", cât și „Fericirea trecută sporește nefericirea prezentă".

Toate emoțiile discutate până acum sunt induse de credințe care sunt (sau ar putea să fie) întreținute la modul certitudinii. Există și emoții – *speranța, frica, iubirea și gelozia* – care presupun într-o manieră esențială credințe întreținute la modul probabilității sau posibilității. Aceste emoții sunt generate de gândul la lucrurile bune sau rele care se pot întâmpla sau nu în viitor ori de stările de lucruri bune sau rele care se pot întâmpla sau nu în prezent.[3] În general, aceste emoții cer ca percepția asupra

---

[1] Am inclus ideea de „non-nemeritat" sub titlul de „meritat". Astfel, când cineva câștigă potul cel mare la loterie, voi spune că acesta a fost meritat într-o oarecare măsură, împotriva utilizării curente a cuvântului.

[2] Cu toate că termenul lui Aristotel pentru această emoție este tradus de obicei prin „indignare", ar trebui să fie clar modul în care aceasta diferă de indignarea carteziană.

[3] Emoțiile pot fi generate, însă probabil că destul de rar, de incertitudinea privitoare la trecut. După prăbușirea fostei Republice Democrate Germane, mulți cetățeni au cerut acces la dosarele securității ca să vadă dacă au fost turnați de prieteni sau de rude. Mai important, unii cetățeni doreau să afle dacă propriile lor eșecuri (faptul că nu au obținut o anumită promovare sau faptul că au fost părăsiți de persoana iubită) se datorează incapacității personale sau faptului că au fost ostracizați de ceilalți la îndemnul securității statului.

evenimentului sau stării de lucruri în cauză să fie că acestea sunt mai mult decât conceptibile, adică trebuie să existe o șansă care nu poate fi neglijată sau o „istorisire cauzală de sus în jos" că evenimentul sau starea de lucruri se pot întâmpla. Gândul că voi câștiga potul cel mare la loterie poate genera speranță, dar nu și gândul „de jos în sus" al primirii unui cadou substanțial de la un milionar necunoscut. Aceste emoții par a presupune, de asemenea, că evenimentul sau starea de lucruri sunt departe de a fi gândite drept certe. Dacă *știu* că urmează să fiu executat, simt mai degrabă disperare decât frică. Conform lui Stendhal, dragostea se ofilește și atunci când afli că este împărtășită, și atunci când afli că nu este. Conform lui La Rochefoucauld, gelozia dispare în momentul în care *știi* că persoana iubită este îndrăgostită de altcineva.

Unele emoții sunt generate de gânduri *contrafactuale* despre ceea ce s-ar fi putut întâmpla sau despre ceea am fi putut face. Dezamăgirea este o emoție care apare când un eveniment despre care se spera că va fi pozitiv nu se materializează.[1] Regretul este o emoție care apare când ne dăm seama că am fi putut face să se întâmple un eveniment despre care se spera că va fi pozitiv dacă am fi ales altfel. Corespondenții pozitivi ai acestor emoții (cauzați de faptul că anumite evenimente negative nu se întâmplă) sunt uneori numiți *exaltare* și, respectiv, *jubilare*. (În limbajul obișnuit, cele două sunt gândite de obicei ca *ușurare*.) În timp ce dezamăgirea și exaltarea presupun compararea unor rezultate diferite cauzate de stări diferite ale lumii, care corespund unor alegeri diferite, regretul și jubilarea presupun comparații cauzate de alegeri diferite în cadrul unei singure stări. În unele cazuri, evenimente negative pot fi atribuite oricăreia dintre surse. Dacă ploaia mă udă în drum spre serviciu, pot atribui acest lucru fie unui eveniment meteorologic întâmplător, fie faptului că nu mi-am luat umbrela. Cu toate că s-ar putea să prefer prima încadrare, această mostră de gândire dezideravă ar putea fi supusă unor constrângeri din partea realității (capitolul 7) în cazul în care abia ascultasem prognoza meteo înainte să ies din casă.

Emoțiile și fericirea

Rolul emoțiilor în producerea fericirii (sau nefericirii) sugerează ideea unui „produs național brut al fericirii". Măsurile obișnuite ale performanței economice sunt, firește, mult mai obiective. Însă, în cele din urmă, nu ne interesează obiectivitatea în sensul capacității fizice de măsurare. Motivul pentru care vrem să știm ceva despre rezultatele unei economii constă în faptul că aceste rezultate contribuie la bunăstarea *subiectivă* sau la fericire. Mai mult, fericirea poate avea surse care nu pot fi măsurate în vreo manieră cantitativă obiectivă. În 1994, anul în care Norvegia a fost gazda Jocurilor Olimpice de Iarnă, țara a trebuit să construiască stadioane noi pentru a face

[1] Ratările la muchie generează emoții mai puternice; astfel, medaliații cu argint ai Jocurilor Olimpice susțin că încearcă o bucurie mai mică decât medaliații cu bronz.

față evenimentelor, precum și spații de găzduire a participanților, costurile acestui demers fiind considerabile. În categoria profitului am putea include banii cheltuiți de turiști și de spectatorii evenimentelor, cât și venitul pe care construcțiile în sine îl vor genera pe termen lung. Economiștii care s-au ocupat de aceste calcule consideră că, per total, investiția a fost o pierdere. Însă eu sunt absolut sigur că, dacă includem beneficiile emoționale ale populației, Jocurile au creat un imens surplus (cu toate că nu pot demonstra acest lucru). Numărul neașteptat de mare de medaliați cu aur din Norvegia a creat o stare de euforie colectivă, al cărei impact a fost uriaș, *tocmai pentru că* victoriile fuseseră atât de neașteptate. Numărul „obiectiv" de victorii și-a datorat o mare parte a impactului elementului de surpriză la nivel subiectiv.[1] Ceva mai recent, victoria echipei de fotbal la Campionatul Mondial din 1998, precum și înfrângerea din 2002 au generat sentimente de euforie și de descurajare, care și-au datorat intensitatea caracterului lor de surpriză.

În general, este greu de făcut o comparație între componenta emoțională a bunăstării sau a stării de bine și alte componente. Faptul că emoțiile pozitive, mai ales când sunt foarte intense, contribuie mai mult la starea de fericire decât simpla bunăstare hedonistă nu dovedește nimic dacă nu știm cât de des survin aceste episoade intense. De asemenea, nu înțelegem dacă și în ce măsură predispoziția pentru cote emoționale maxime merge mână în mână cu predispoziția pentru cote emoționale minime. Dacă lucrurile stau astfel, ne putem întreba dacă o viață de mulțumire stabilă este mai fericită per ansamblu decât una în care alternează momentele de euforie și cele de disforie. După cum observă Montaigne, răspunsul depinde de oportunitățile oferite de mediul înconjurător: „Sunt de acord că avantajul de a ni se îngheța și toci simțurile după ce gustăm din durerile nefaste trebuie să aducă dezavantajul că devenim mai puțin în măsură să apreciem bucuriile plăcerilor bune. Însă mizeria condiției umane presupune că există mai puține lucruri de care ne putem bucura și mult mai multe pe care trebuie să le izgonim." Idealul anihilării emoțiilor, pe care îl regăsim în multe filosofii antice, în special în stoicism și budism, a apărut în societăți unde mediul oferea probabil mai multe ocazii emoțiilor cu valență negativă. Pentru că scria în timpul războaielor religioase care au devastat Franța, Montaigne se afla, probabil, în aceeași situație.

Emoție și acțiune

Legătura dintre emoție și acțiune este aceea a unei tendințe spre acțiune (sau disponibilitate pentru acțiune). De asemenea, putem concepe tendința spre acțiune ca pe

[1] Să presupunem că probabilitatea anterioară a unei victorii este p și că satisfacția obținută dintr-o victorie este proporțională cu $1/p$ (întrucât surpriza este mai mare când p este mic). În acest model special al surprizei, satisfacția așteptată a victoriei este independentă de probabilitatea victoriei. În general, impactul surprizei va fi mult mai complex.

o preferință temporară. Fiecare emoție importantă pare să fie asociată cu una (sau mai multe) astfel de tendințe (vezi Tabelul 8.1).

Cu toate că mânia și indignarea carteziană determină aceeași tendință spre acțiune, cea asociată cu mânia este mai puternică. Există experimente care pun în evidență faptul că subiecții sunt dispuși să suporte un cost mai mare pentru a răni pe cineva care i-a rănit pe ei, decât pentru a răni pe cineva care a rănit un terț. După sfârșitul celui de-al Doilea Război Mondial, americanii erau mult mai dispuși să-i pedepsească pe naziștii care i-au tratat prost pe prizonierii de război americani decât pe aceia responsabili pentru Holocaust. O excepție care confirmă regula este reprezentată de membrii evrei ai administrației Roosevelt.

Emoțiile de mânie, vină, dispreț și rușine au legături apropiate cu normele morale și sociale. Aceia care încalcă normele suferă vină sau rușine, în timp ce aceia care sunt martorii încălcărilor simt mânie sau dispreț. Structurile acestor relații diferă, după cum reiese din Figura 8.1.

Normele sociale, pe care urmează să le discut în capitolul 22, sunt mijlocite de expunerea la ceilalți. De aceea, sinuciderile menționate la începutul acestui capitol au avut loc doar când acțiunile rușinoase au fost făcute publice. După cum am susținut în capitolul 5, normele morale sunt diferite din acest punct de vedere.[1]

TABEL 8.1

Emoție	*Tendință spre acțiune*
Mânie sau indignare carteziană	Cauzează suferință obiectului emoției.
Ură	Cauzează încetarea existenței obiectului urii.
Dispreț	Ostracizare, evitare.
Rușine	„Intră în pământ", fugă, sinucidere.
Vină	Mărturisire, încercarea de a repara lucrurile, auto-pedepsire.
Invidie	Distrugerea obiectului invidiat sau a proprietarului acestuia.
Frică	Fugă, luptă.
Iubire	Apropiere de celălalt, atingerea lui; ajutor acordat celuilalt; a-i face pe plac celuilalt.
Milă	Consolarea sau alinarea nefericirii celuilalt.
Recunoștință	A-l ajuta pe celălalt.

[1] Bănuiesc că violarea normelor cvasi-morale declanșează aceleași emoții ca și violarea normelor morale, dar intuiția mea nu este foarte puternică.

BCU IASI/CENTRUL UNIVERSITAR

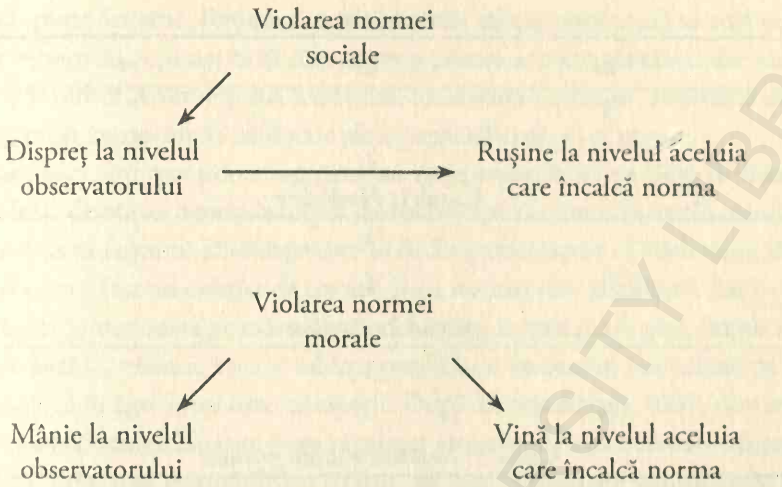

FIGURA 8.1

Unele tendințe spre acțiune par a urmări să „readucă echilibrul moral al universului". A le face rău acelora care ți-au făcut rău și a-i ajuta pe aceia care te-au ajutat sunt moduri de a *echilibra lucrurile*. În anumite cazuri, acest lucru este adevărat. Însă teoria perspectivei (capitolul 12) presupune că „Doi ochi pentru un ochi" poate fi o descriere mai bună a tendinței spre acțiune a mâniei decât „Un ochi pentru un ochi".[1] Cu toate că multe acte de recunoștință pot fi făcute doar pentru a te achita de o datorie, altele pot să își aibă originea într-un sentiment autentic de bunăvoință față de acela care ți-a făcut un bine. Viziunea despre echilibrul moral este mai convingătoare în cazurile de vină, pentru care tendința spre acțiunea de a repara lucrurile are în mod explicit un caracter restaurator. Mai mult, când agentul nu poate face nimic pentru a șterge răul făcut, poate reface echilibrul făcându-și sieși rău într-o măsură egală. Dacă nu mi-am plătit taxele și descopăr că Fiscul nu acceptă donații anonime de bani pentru suma pe care o datorez, pot restaura echilibrul dând foc la bani.

Tendințele emoționale spre acțiune nu determină doar o dorință de a acționa. Ele determină și dorința de *a acționa mai devreme mai degrabă decât mai târziu*. Pentru a contextualiza această idee, voi face o distincție între *nerăbdare* și *presiune*. Definesc

[1] Astfel, ca răspuns pentru uciderea a 365 de musulmani libanezi, o femeie libaneză a spus că, „în acest moment, vreau ca [poliția musulmană] (...) să intre în birouri și să ucidă primii șapte sute treizeci de creștini fără apărare pe care pot pune mâna". În *Războiul Civil din Franța*, Marx citează o scrisoare publicată într-un ziar și aparținând unui preot luat ostatic de revoluționari, în care acesta spune că „a fost luată decizia de a ucide câte doi ostatici pentru fiecare nouă execuție" (a forțelor guvernamentale). În cele din urmă, și Seneca observă că „un rău căruia nu i se răspunde cu un rău mai mare nu a fost răzbunat".

TABEL 8.2

t1	*t2*	*t3*	*t4*
A	3		
B		5	
Cazul 1: *Nerăbdare*			
A		3	
	B	4	
Cazul 2: *Presiune*			
A		3	
	B		6
Cazul 3: *Nerăbdare și/sau presiune*			

nerăbdarea ca o preferință pentru o recompensă venită mai devreme în favoarea uneia venită mai târziu, cu alte cuvinte, avem de-a face cu o anumită depreciere a timpului. După cum am observat în capitolul 6, emoțiile pot face ca un agent să acorde o importanță mai mică unor consecințe îndepărtate în timp ale unei acțiuni prezente. Pentru o ilustrare a distincției, vezi Tabelul 8.2.

În fiecare caz, agentul poate alege una, și doar una, dintre acțiuni, A sau B. În cazul 1, aceste opțiuni sunt valabile în același timp, în cazurile 2 și 3 sunt valabile la momente succesive. În cazul 2, recompensele (a căror magnitudine este indicată prin numere) survin în același moment de mai târziu, pe când în cazurile 1 și 3 în momente succesive mai târzii. Să presupunem că, într-o stare lipsită de emoție, agentul alege B în toate cazurile, dar într-o stare emoțională alege A. În cazul 1, alegerea lui A se datorează nerăbdării induse emoțional. În cazul 2, se datorează presiunii induse emoțional. În cazul 3, se poate datora ambelor sau unei interacțiuni între cele două.

Nerăbdarea, o chestiune mult discutată, a fost un subiect central al capitolului 6. Cu toate că presiunea este mai puțin discutată, cred că poate fi foarte importantă. În mod particular, presiunea emoțiilor ne furnizează unul dintre mecanismele prin care acestea pot influența formarea credințelor. După cum vom vedea în capitolul 11, formarea de credințe raționale cere o strângere optimă de informații. În loc să urmeze probele deja disponibile, un agent rațional va aduna probe suplimentare înainte să acționeze dacă decizia care trebuie luată este suficient de importantă iar costul așteptării suficient de mic. Emoțiile presante sunt de multe ori declanșate în situații în care costul așteptării *este* mare, adică în fața unui pericol fizic acut. În astfel de cazuri, este esențial să acționezi rapid, fără să stai pe gânduri pentru a descoperi mai multe lucruri privitoare la situație. Dar când o decizie importantă poate fi îmbunătățită prin așteptare, o dorință indusă de emoție în vederea acțiunii imediate poate fi dăunătoare.

După cum spune Seneca, „Rațiunea ascultă ambele tabere, apoi caută să amâne acțiunea, chiar propria sa acțiune, ca să câștige timp pentru a cerne adevărul; dar mânia se grăbește". Proverbul „Cine-n pripă se însoară, toată viața-i e amară" sugerează atât forța emoției, cât și consecințele nefericite ale incapacității de a i te opune.[1]

Mai devreme, am menționat faptul că nu toate emoțiile au un timp de înjumătățire scurt. Însă, de obicei, emoțiile dispar destul de rapid în timp. În unele cazuri, acest lucru se datorează faptului că situația care le declanșează dispare. Odată ajuns în siguranță după ce am fost amenințat de un urs, frica nu mai este justificată. Însă, cel mai adesea, emoția se atenuează pe măsură ce amintirea ei se atenuează, prin simpla trecere a timpului. Mânia, rușinea, vina și iubirea persistă rar cu aceeași intensitate pe care o au la momentul în care izbucnesc ca emoții. După 11 septembrie 2001, de exemplu, numărul de tineri americani care și-au exprimat interesul de a face armata a înregistrat o creștere cu 50%, însă recrutările nu au fost, de fapt, influențate semnificativ. Aceste fapte confirmă ipoteza că valul inițial de interes se datora emoției, care s-a atenuat în lunile următoare ale procesului de recrutare. În rândul femeilor tinere nu a fost înregistrată vreo modificare semnificativă a interesului pentru a se înrola în armată, fapt care nu are o explicație evidentă.

Acestea fiind spuse, oamenii sunt adesea incapabili să anticipeze diminuarea emoțiilor lor. Când sunt cuprinși de emoție, aceștia pot crede, în mod greșit, că sentimentul va dura veșnic sau pot chiar pierde perspectiva viitorului. Dacă indivizii sinucigași la care m-am referit mai devreme ar fi știut că emoția lor se va diminua cu trecerea timpului (la fel și disprețul observatorilor), poate că nu s-ar fi sinucis. Dacă tinerele cupluri ar ști că iubirea lor nu va dura veșnic, poate că ar fi mai puțin dispuse să își asume anumite angajamente sau să intre într-o „căsătorie-legământ", din care este mult mai dificil să iasă.

Voi încheia acest argument prin indicarea unei interacțiuni între două fenomene induse de emoție: inversarea preferințelor și formarea de credințe tulburi. Într-o perspectivă optimistă, aceste două fenomene se pot anula una pe cealaltă întrucât, din cauza inversării preferințelor, o persoană vrea să acționeze împotriva propriei gândiri calme și reflexive cu privire la ce e bine, dar, din cauza credinței tulburi, nu poate să materializeze intenția. Cel mai adesea, cred, cele două se vor întări reciproc. Răzbunarea este un exemplu. Riscul este minim dacă nu mă răzbun pentru un afront este,

---

[1] Reducerea libertăților cetățenești, legiferată de mai multe guverne din Occident după 11 septembrie 2001, este un bun studiu de caz. Oare avem de-a face cu o nevoie de a acționa imediat în prezența unui pericol iminent sau cu o reacție panicardă care, făcând aceste guverne să pară chiar mai odioase în ochii dușmanilor lor, mai degrabă crește probabilitatea unor atacuri teroriste? Putem face și o comparație cu „psihologia tiraniei", pe care am discutat-o în capitolele anterioare.

mai mare dacă nu mă răzbun, dar aștept o vreme, și este maxim dacă mă răzbun imediat fără să fiu preocupat de riscuri. Montaigne face o observație asemănătoare: „Când pedepsim orice afront care ni s-a adus, filosofia ne învață să evităm mânia, nu pentru a ne diminua răzbunarea, ci (dimpotrivă) pentru ca forța ei să fie mai mare și mai bine țintită; filosofia consideră că emoțiile violente sunt un impediment pentru aceasta." Însă acest lucru presupune să ignorăm paradoxul conform căruia, dacă nu simțim emoție, nu mai vrem să ne răzbunăm, iar, dacă o simțim, e posibil să nu fim în stare să ducem la îndeplinire răzbunarea într-un mod eficient.

Emoție și credință

Emoțiile pot influența formarea credințelor atât direct, cât și indirect. Efectul direct produce credințe părtinitoare, efectul indirect produce credințe de proastă calitate. O formă de părtinire este ilustrată în teoria *cristalizării* a lui Stendhal. Originea termenului este următoarea: „La minele de sare din Hallein, de lângă Salzburg, minerii aruncă o ramură înghețată fără frunze în puțurile abandonate. Două sau trei luni mai târziu, prin efectul apelor saturate cu sare care îmbibă ramura, pentru a o lăsa apoi să se usuce după ce acestea se retrag, minerii o găsesc acoperită cu o depunere strălucitoare de cristale. Cea mai mică ramură, nu mai mare de un deget, este încrustată cu o infinitate de cristale scânteietoare și fascinante." Analogia cu dragostea este evidentă: „Din momentul în care începi să fii interesat cu adevărat de o femeie, nu o mai vezi așa *cum e aceasta de fapt*, ci așa cum îți convine să o vezi. Compari iluziile flatante create de acest interes incipient cu diamantele frumoase care ascund o banală ramură fără frunze și care, atenție, sunt percepute prin ochii tânărului îndrăgostit."

Un proverb francez pe care l-am citat și mai devreme spune că *credem cu ușurință ceea ce ne provoacă frică*. Și aceasta este o formă de părtinire. Pe lângă faptul că, în mod natural (chiar și în stări neemoționale), avem tendința de a acorda o importanță excesivă riscurilor cu probabilitate mică (capitolul 12), sentimentele de frică viscerală ne pot face să credem că pericolele sunt mai mari decât sunt de fapt. Când traversăm o pădure în toiul nopții, un sunet sau o mișcare pot declanșa frica, senzație care ne face apoi să interpretăm ca înfricoșătoare alte sunete sau mișcări pe care altădată le-am ignorat. Frica „se hrănește pe sine". Ne putem gândi și la „efectul Othello". Voi mai discuta aceste chestiuni și în capitolul 23.

Presiunea emoțiilor acționează asupra culegerii de informații înainte de formarea credinței, și nu asupra credinței în sine. Rezultatul este o credință de proastă calitate bazată pe o cantitate de informații care nu este cea optimă, și nu o credință părtinitoare în favoarea unei concluzii anume despre care agentul ar dori să fie adevărată. Însă, în practică, cele două mecanisme merg mână în mână și se întăresc reciproc. Agentul își formează inițial o atitudine părtinitoare indusă de emoție, apoi presiunea emoției îl împiedică să adune informații care ar fi putut corecta părtinirea. După cum am

remarcat în capitolul trecut, gândirea dezideративă este supusă, într-o oarecare măsură, unor constrângeri impuse de realitate. Gândirea antidezideративă este constrânsă, de asemenea, în același mod. Prin urmare, dacă agentul ar fi adunat mai multe informații, poate că nu ar fi persistat în a întreține o credință părtinitoare.

Trebuie, totuși, să repet că este posibil ca strângerea informațiilor să nu fie rațională dacă implică niște costuri de oportunitate mari. Dacă stai prea mult pe gânduri încercând să-ți dai seama dacă forma din cale este un băț sau un șarpe, s-ar putea să mori. De asemenea, trebuie adăugat că, în stări neemoționale, există o tendință de a *desconsidera* costurile de oportunitate și de a acorda o mai mare atenție unor cheltuieli mărunte (capitolul 12). În fața unui posibil risc, un agent poate fi convins să nu își asume precauții prea mari din cauza costului lor direct, fără să țină cont de faptul că inacțiunea poate fi, de asemenea, costisitoare. Prin urmare, în anumite circumstanțe presiunea emoției poate aduce o corijare utilă acestei tendințe iraționale. În același timp, repet, presiunea și credințele de proastă calitate creează mai multe probleme decât rezolvă.

Emoții și transmutație

Dată fiind cauza ierarhia normativă a motivațiilor (capitolul 4), oamenii se pot rușina de emoțiile lor. Invidia, de pildă, este o emoție pe care cei mai mulți nu recunosc că o simt. Când Iago spune: „Dacă scapă Cassio/E-n viața lui ceva măreț de-a pururi, și-mi face umbră"[1], acesta este neobișnuit, poate implauzibil, de sincer. Reacția cea mai obișnuită la vederea invidiei proprii este o ridicare din umeri la nivel mental. Senzația este una de inferioritate dureroasă, urmată de o dorință destructivă, apoi lucrurile revin la normal. Uneori, însă, emoția este atât de puternică, încât nu poate fi ignorată. Dar, în același timp, ea nu poate fi recunoscută. Soluția conflictului este ca invidia să fie transmutată în indignare îndreptățită prin intermediul unei rescrieri convenabile a scenariului. Îmi pot spune mie însumi o poveste în care celălalt a obținut obiectul invidiat prin mijloace imorale și ilegitime și, poate, în dauna mea, preschimbând astfel invidia în indignare aristotelică sau în mânie.

În capitolul 7 am discutat formarea motivată a credințelor. După cum arată Figura 8.2, acest fenomen poate fi inclus în procesul de *motivare motivată*, pe care l-am discutat în capitolul 4. Pentru ca un agent să fie în măsură să adopte o motivație de care nu îi este rușine poate fi necesară o rescriere la nivel cognitiv.

Această metamorfoză a invidiei în indignare îndreptățită poate să apară și într-o manieră directă. Invidia nu este doar rușinoasă, ci și dureroasă. Credința că altcineva a avut succes acolo unde aș fi putut reuși eu dacă aș fi depus un efort mai mare poate

[1] *Othello*, Actul V, Scena 1, în traducerea lui Ion Vinea. (N. red.)

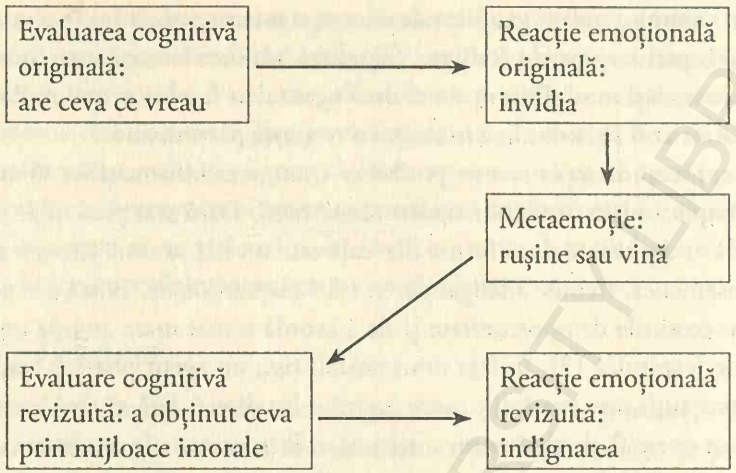

FIGURA 8.2

fi extrem de neplăcută. Pentru a atenua durerea, pot adopta o istorisire asemănătoare despre cauzele nerespectabile ale succesului celuilalt. Vina poate fi, de asemenea, foarte greu de suportat, mai ales pentru o persoană arogantă. Indivizii aroganți vor putea foarte ușor să inventeze o istorisire prin care să dea vina pe victima lor, și nu pe ei înșiși: „Pe aceia pe care îi rănesc îi și urăsc", spunea Seneca. Acest model pare destul de obișnuit și poate fi întâlnit, de pildă, în cazul prinților din Renaștere, în cazul tiranilor antici și al altor indivizi autoritari.

Preocuparea pentru imaginea proprie poate cauza evitarea regretului. Pentru unii indivizi, să admită față de sine că au făcut o greșeală este atât de dureros, încât aceștia s-ar angaja mai degrabă într-o activitate foarte neprofitabilă decât să-și accepte eșecul, după cum ar dicta gândirea rațională. Aceasta este „eroarea costurilor investite", cunoscută și sub numele de „eroarea vietnameză" sau „eroarea Concorde." Voi aborda chestiunea în capitolul 12.

Cultura și emoțiile

Sunt toate emoțiile universale? Dacă nu, există oare emoții universale? Răspund cu un „da" hotărât la ultima întrebare și cu un „da" cu jumătate de gură la prima întrebare.

Pare destul de clar că anumite emoții sunt universale. Există câteva emoții – fericire, surprindere, frică, tristețe, dezgust și mânie – a căror expresie facială este recunoscută de oameni din culturi diferite. În cazul în care credeți, alături de mine, că normele sociale există în toate societățile, emoțiile care le susțin – dezgustul și rușinea – trebuie să fie și ele universale. Ne putem imagina o societate în care oamenii simt emoția mâniei atunci când sunt ofensați, dar nu simt indignare (carteziană) când

observă ofense aduse unor terți. Îmi este greu să îmi imaginez o astfel de societate, deși poate mă înșel. Dacă iubirea este universală (vezi discuția de mai jos), de ce nu ar fi și gelozia universală?

Se spune că japonezii au o emoție numită *amae* (echivalentă, în mare, unei stări de neajutorare și dorinței de a fi iubit), care nu există în alte societăți. S-a susținut, de asemenea, că societatea greacă era o „cultură a rușinii", diferită substanțial de moderna „cultură a vinovăției", că dragostea romantică este o invenție modernă și că sentimentul plictisului (dacă îl putem considera o emoție) este de dată recentă. Nu putem, totuși, exclude faptul că emoțiile așa-zis absente poate existau, dar nu au fost conceptualizate de membrii societăților în discuție. O emoție poate fi recunoscută ca atare de un observator extern, dar poate să nu fie recunoscută de membrii societății respective. În Tahiti, un bărbat părăsit de prietena sa va manifesta toate simptomele comportamentale ale tristeții, dar va spune doar că este „obosit". În Occident, *conceptul* de iubire romantică este relativ recent, datând din vremea trubadurilor. Înainte de acea perioadă, exista doar „senzualitate plăcută sau nebunie". Însă este posibil și, din punctul meu de vedere, foarte probabil ca *experiența* iubirii romantice să fi existat chiar și atunci când societatea nu avea un concept pentru această emoție. Indivizii pot fi îndrăgostiți fără să conștientizeze acest lucru și, în același timp, emoțiile lor pot fi evidente pentru observatori, fie aceștia din societatea respectivă sau din alta. Grecii antici manifestau mai multe reacții înrudite cu vina – mânia, iertarea și compensațiile – care indică prezența emoției, chiar dacă nu aveau un cuvânt care să o descrie. Modul în care oamenii se gândesc la emoții poate fi specific unei anumite culturi, chiar dacă emoțiile în sine nu sunt.

Ar trebui să adăugăm, totuși, că, atunci când o anumită emoție nu este conceptualizată explicit, aceasta poate avea mai puține manifestări comportamentale. La Rochefoucauld scria că „unii oameni nu ar ajunge niciodată să se îndrăgostească dacă n-ar fi auzit vreodată de iubire". De asemenea, vina poate că este mai des întâlnită în societățile în care oamenilor li se spune de la o vârstă fragedă că ar trebui să se simtă vinovați în cutare sau cutare moment.

◈◈◈

Notă bibliografică

Cea mai bună carte despre emoții este scrisă de N. Frijda, *The Emotions* (Cambridge University Press, 1986). Am folosit destul de mult această sursă în propria mea carte *Alchemies of the Mind* (cu subtitlul *Rationality and the Emotions*) (Cambridge University Press, 1999), unde cititorul poate găsi și alte trimiteri la conceptele discutate aici. Ideea de presiune, ca fiind diferită de nerăbdare, reprezintă o completare a cadrului

conceptual din acea carte. Discuții extinse despre rolul emoțiilor în explicarea comportamentului se pot găsi în R. Petersen, *Understanding Ethnic Violence: Fear, Hatred, and Resentment in Twentieth-Century Eastern Europe* (Cambridge University Press, 2002), și în capitolul 8 din cartea mea *Closing the Books: Transitional Justice in Historical Perspective* (Cambridge University Press, 2004).

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

III
ACȚIUNEA

Cu toate că voi folosi „acțiune", „comportament", „decizie" și „alegere" ca termeni sinonimi, uneori este util să îi distingem. Cea mai cuprinzătoare categorie este *comportamentul*, înțeles ca orice mișcare a corpului a cărei origine este internă agentului, și nu externă (ca atunci când acesta este purtat de o alunecare de teren). Acțiunea este comportamentul intenționat, cauzat de dorințele și credințele agentului. Astfel, comportamentele reflexe nu sunt acțiuni; faptul de a avea erecție nu este o acțiune (dar poate fi indus de o acțiune, cum ar fi administrarea de Viagra); faptul de a adormi nu este o acțiune (dar poate fi indus de administrarea unui somnifer). O acțiune poate fi precedată sau nu de o *decizie* conștientă. Când conduc spre serviciu pe drumul meu obișnuit, nu decid conștient să fac stânga sau dreapta, cu toate că fiecare acțiune este intenționată și orientată spre un anumit scop. Însă, când conduc spre serviciu pentru prima sau a doua oară, acțiunile sunt precedate de decizii explicite. De fapt, sunt precedate de o *alegere* explicită între rute alternative. Cu toate că toate alegerile sunt decizii, nu toate deciziile sunt alegeri. Când decid să iau de pe masă cartea pe care o citesc de ceva vreme, nu trebuie să am o alternativă explicită în minte. Văd cartea pe masă; observarea ei îmi aduce aminte de faptul că îmi face plăcere să o citesc și decid să o ridic. Nu este implicată nicio alegere.

Cea mai importantă trăsătură a acestui cadru conceptual este aceea că *nu toate deciziile conduc la acțiune*. Cineva ar putea decide să nu facă un anumit lucru, să nu salveze o persoană de la înec dacă o astfel de intervenție este mult prea riscantă pentru sine. Dacă persoana se îneacă și nu este implicat un terț, nu am nicio responsabilitate cauzală pentru rezultat. Poate că am o responsabilitate morală, iar în unele țări chiar legală, dar aceasta este o altă problemă.[1] Dar să presupunem că există o terță persoană

[1] În Statele Unite nu există vreo datorie să fii bunul samaritean, cu excepția unor circumstanțe foarte precis delimitate. În Europa continentală, „non-asistarea unei persoane aflate în pericol" poate fi pedepsită cu severitate dacă riscul înfruntat de samaritean pentru a-și oferi ajutorul este mult mai mic în comparație cu pericolul cu care se confruntă persoana care are

BOSTON UNIVERSITY LIBRARY

prezentă sau, ca în cazul lui Kitty Genovese, mai multe persoane implicate. Dacă terța persoană observă că sunt în măsură să ajut persoana care se înecă, dar nu o fac, aceasta poate trage concluzia, perfect rezonabilă, că situația este mult mai puțin serioasă decât ar putea să pară și, prin urmare, să nu sară în ajutor. În acest caz, decizia mea de a nu face nimic va fi cauzat decizia altei persoane de a nu face nimic. Astfel, deciziile pot avea eficiență cauzală chiar atunci când nu generează o acțiune.

În bună măsură, Partea a III-a este organizată în jurul teoriei alegerii raționale. După cum am explicat în Introducere, am ajuns să fiu ceva mai sceptic decât înainte cu privire la explicațiile acțiunii (sau inacțiunii) bazate pe alegerea rațională. Însă, cu toate că o mare parte a comportamentului este irațională, într-un fel sau altul, există un sens în care raționalitatea rămâne primară. Ființele umane vor să fie raționale. Nu ne mândrim cu derapajele noastre în iraționalitate. Mai degrabă, încercăm să le evităm sau să le corectăm, aceasta în cazul în care aroganța noastră nu ne împiedică să le recunoaștem.

---

nevoie de ajutor. Unii specialiști în drept americani susțin că legea americană este mult mai eficientă din moment ce o datorie generală de a ajuta creează un stimulent pentru salvatorii potențiali de a evita acele locuri în care s-ar putea să fie nevoie de ei din cauza amenințării impuse de responsabilitate. S-ar putea (dar mai degrabă nu) ca lucrurile să stea astfel; ce pare sigur este că sistemul american nu își are originea în motive de eficiență și nici nu a fost menținut pentru proprietățile sale legate de eficiență.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 9

DORINȚE ȘI OPORTUNITĂȚI

A face tot posibilul

Uneori, caracterizăm un anumit comportament prin remarca „A făcut tot posibilul". Dacă analizăm propoziția, remarcăm prezența a două elemente: dorințele și oportunitățile. Dorințele definesc ceea ce este „cel mai bun lucru" pentru agent. Oportunitățile sunt opțiunile sau mijloacele din care agentul „poate" alege. Această caracterizare poate servi și drept *explicație* rudimentară, pe modelul alegerii raționale, a comportamentului în cauză. Dacă întrebăm „De ce a făcut acest lucru?", răspunsul „A făcut tot posibilul" poate fi de ajuns. În multe cazuri, este nevoie de mai mult pentru a produce o explicație satisfăcătoare din perspectiva alegerii raționale. În particular, poate fi nevoie să recurgem la *credințele* agentului, și nu doar la dorințe și oportunități. Aceste aspecte mai complicate vor forma obiectul capitolului 11. Aici voi discuta cât de departe putem merge folosind cadrul simplu al dorințelor și oportunităților.[1] Sugerez, de asemenea, că nu avem de-a face cu un cadru atât de simplu cum pare la prima vedere din moment ce dorințele și oportunitățile nu sunt întotdeauna (după cum se crede uneori) independente unele de altele.

Există o manieră diferită, dar echivalentă, în care putem privi situația. În înțelegerea comportamentului putem începe prin a lua în considerare toate posibilitățile abstracte de acțiune în care se poate implica un individ. Acțiunea manifestă poate fi considerată drept rezultatul a două opțiuni succesive de filtrare. Primul filtru îl constituie toate *constrângerile* – fizice, economice, legale și altele – cu care se confruntă agentul. Acțiunile coerente cu toate aceste constrângeri constituie setul de oportunități.[2] Al doilea filtru este un mecanism ce determină care dintre acțiunile din cadrul

---

[1] Procedând astfel adopt premisa implicită că setul de opțiuni pe care agentul *crede* că le are la dispoziție coincid cu setul „obiectiv" de oportunități.

[2] Când vorbesc despre constrângeri legale nu mă refer la efectul legilor de a face anumite acțiuni mai costisitoare decât altele (acest lucru ține de filtrul secundar), ci la efectul lor de a

de oportunități va fi dusă la îndeplinire. Presupun că agentul alege acțiunea care va avea cele mai bune consecințe, așa cum sunt acestea evaluate prin dorințele sau preferințele sale. În capitolele următoare voi lua în considerare și alte mecanisme secundare de filtrare.

Abordarea prin filtre sugerează următoarea întrebare: ce se întâmplă în cazul în care constrângerile sunt atât de puternice încât filtrul secundar nu mai are nimic de filtrat? Se poate întâmpla să avem de-a face cu constrângeri care determină în mod unic un anumit curs al acțiunii consistent cu acestea? Atât bogații, cât și săracii au oportunitatea de a dormi sub podurile din Paris, dar săracii poate că nu au altă oportunitate.[1] Pentru un consumator sărac, constrângerile de natură economică și calorică pot determina împreună un unic coș de bunuri.[2] Cei care apără ideea de structuralism în științele sociale pot fi interpretați ca susținând ideea că, în mod obișnuit, constrângerile sunt atât de puternice încât lasă loc foarte mic sau deloc pentru alegeri.[3] De ce ar sta lucrurile astfel rămâne, însă, un mister. Nu putem susține, de pildă, că bogații și cei puternici se asigură că săracii și oprimații nu au altă opțiune decât aceea de a-i servi întrucât presupunem astfel că măcar bogații și cei puternici au cel puțin o alegere la dispoziție.

În anumite cazuri, abordarea prin dorințe și oportunități este incompletă. Pot avea oportunitatea să fac o acțiune care va realiza cel mai bine dorința pe care o am, cum ar fi alegerea răspunsului corect la un test sau ochirea țintei într-un concurs de tir, dar nu pot avea și *abilitatea* de a identifica acțiunea care va aduce rezultatul dorit. În unele cazuri, această inabilitate poate fi pusă pe seama unei etape anterioare, în care agentului îi lipsește atât oportunitatea, cât și dorința de a dobândi respectiva abilitate. În alte cazuri, inabilitatea este consecința unor constrângeri dure de natură psihofiziologică. Pot avea atât dorința, cât și oportunitatea de a alege acțiunea care maximizează bunăstarea mea pe termen lung, dar nu și abilitatea de a hotărî pe loc ce acțiune

---

face anumite acțiuni posibile sau imposibile. Nu pot vota decât în ziua alegerilor și nu mă pot căsători decât în locurile ratificate de lege.

[1] Dacă descriem setul de oportunități mai bine, săracii pot avea la dispoziție alegerea sub care pod să doarmă. Acesta este un aspect general: pentru orice descriere a opțiunilor putem fi în măsură să specificăm o situație în care agentul are o singură opțiune fezabilă, dar pentru orice situație putem găsi o descriere în care există mai mult de una.

[2] În mod obișnuit, există mai multe strategii pentru a supraviețui la nivel de subzistență. Spre deosebire de alegerea podului sub care poți dormi, acestea diferă într-un mod destul de semnificativ. Una dintre erorile teoriei valorii-muncă a lui Marx este incapacitatea acestuia de a înțelege acest fapt.

[3] O altă concepție asupra structuralismului, fără legătură cu cea de aici, a fost menționată în capitolul 1.

anume poate aduce acest rezultat. Când economiștii și cercetătorii din științele politice care gândesc matematic încearcă să identifice comportamentul optim într-o situație dată, aceștia au nevoie de multe pagini pline de formule complicate pentru a ajunge la un rezultat. Pentru unii indivizi, oricâte exerciții ar face, aceste calcule se află peste puterile lor sau sunt suficient de complicate pentru a implica mai mult efort decât beneficiile aferente sarcinii. Timpul presupus de astfel de calcule poate fi, de asemenea, destul de îndelungat pentru a fi disponibil.

Explicațiile bazate pe oportunități

Chiar și atunci când comportamentul este rezultatul comun al dorințelor și oportunităților, *varianța* în comportament de-a lungul timpului poate fi explicată în bună măsură prin apel la oportunități. Consumul de alcool este, în general, determinat atât de intensitatea dorinței oamenilor de a bea, comparată cu alte dorințe, cât și de ceea ce își pot permite. De pildă, în vreme de război, când prețurile la alcool cresc brusc, consumul de alcool scade drastic.

O explicație poate fi dată în termeni de curbe de indiferență (Figura 9.1). Să presupunem că un consumator trebuie să își împartă veniturile între alcool și un pachet de bunuri de consum obișnuite. Prețurile relative și veniturile persoanei sunt la început de așa natură încât aceasta are în față setul de oportunități delimitat de triunghiul

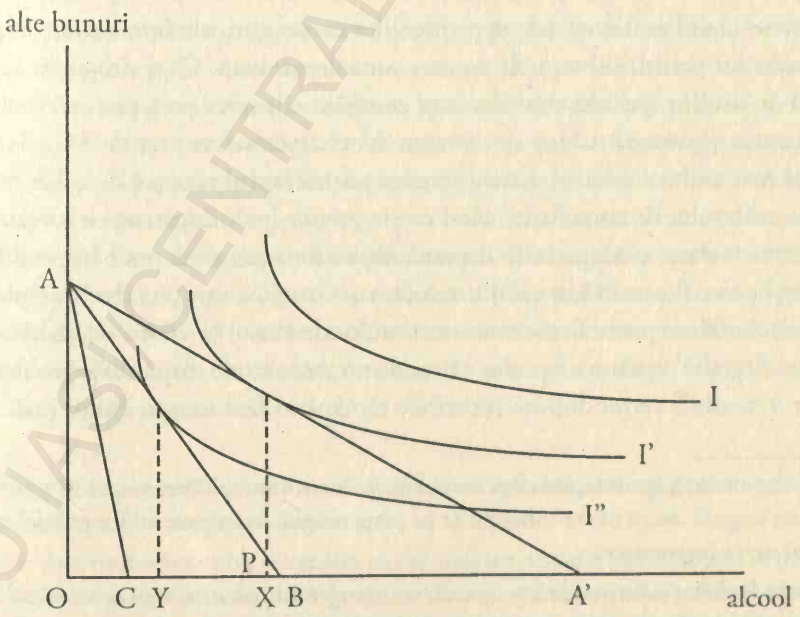

FIGURA 9.1

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

OAA'. Presupunând că persoana își cheltuiește întregul venit, ne putem limita la linia bugetului AA'.[1] Intensitatea dorinței sale pentru alcool *versus* pachetul de consum este prezentată sub forma *curbelor de indiferență* I, I' și I". Termenul reflectă ideea că un consumator este indiferent față de toate combinațiile de alcool și alte bunuri care se află pe oricare dintre curbele date, preferând în același timp combinațiile de pe o curbă superioară oricărei alteia de pe o curbă inferioară.[2] Pentru a alege cea mai bună dintre opțiunile disponibile, consumatorul trebuie să aleagă punctul aflat pe linia bugetului care este tangent unei curbe de indiferență, din moment ce aceasta va fi curba superioară tuturor curbelor care includ o combinație pe care persoana și-o poate permite. În Figura 9.1, aceasta ne dă consumul de alcool OX.

Să presupunem acum că prețul la alcool crește, astfel încât consumatorul are în față linia de buget AB. Din moment ce punctul de tangență s-a mutat la stânga, consumatorul va consuma acum OY. Putem face același raționament dacă o nouă creștere de preț va muta linia bugetului la AC. Însă, chiar dacă nu știm nimic despre forma curbelor de indiferență, putem face predicția că, într-o astfel de situație, consumatorul nu va consuma mai mult de OC, caz în care ar trebui să cheltuiască tot venitul pe alcool. Setul de oportunități poate să ne dea prin sine o explicație substanțială a variației de-a lungul timpului. De fapt, al doilea filtru poate fi orice – îmbunătățirea comportamentului, o poftă irezistibilă după alcool, obișnuința etc. –, consumatorul tot va fi sever restricționat de primul filtru.

Am ales acest exemplu particular pentru a discuta chestiunea așa-ziselor dorințe „irezistibile", cum ar fi dorințele dependenților de droguri, ale fumătorilor înrăiți sau ale alcoolicilor pentru substanțele de care sunt dependenți. Oare drogurile sunt mai degrabă ca insulina, pe care un diabetic ar cumpăra-o la orice preț, sau mai mult ca zahărul, care înregistrează scăderi de consum odată cu creșterea prețului? Ca dovadă că seamănă mai mult cu zahărul se menționează adesea faptul că avem de a face cu o scădere a consumului de stupefiante când crește prețul. Însă, după cum am văzut, acest lucru se poate datora incapacității dependentului de a consuma peste bugetul său. (Și diabeticul poate fi pus în imposibilitatea de a-și cumpăra insulina dacă prețul acesteia crește.) Astfel, se pare că scăderea consumului de alcool în vreme de război trebuie pusă mai degrabă pe seama faptului că acesta nu este atât de disponibil, problema dorințelor irezistibile *versus* dorințe rezistibile rămânând fără soluție. Știm, însă, pe alte

---

[1] Cu alte cuvinte, ignor faptul că persoana poate lucra ore suplimentare, că își poate prepara băutura singură sau că poate cumpăra de pe piața neagră. În aplicarea unor politici, astfel de chestiuni devin importante.

[2] Forma curbelor corespunde faptului că, cu cât agentul consumă mai mult alcool, cu atât are nevoie să compenseze (și să rămână la același nivel de bunăstare) pentru o anumită reducere a pachetului de bunuri de consum.

temeiuri că există o sensibilitate la preț în cazul consumului de alcool. Chiar și atunci când consumatorii își permit să mențină nivelul anterior de consum la prețuri mai mari, aceștia nu o fac.

Uneori, se oferă un alt argument pentru a pune în evidență faptul că oportunitățile au o putere explicativă mai mare decât dorințele. Unii economiști susțin că toți indivizii au, esențialmente, aceleași dorințe și preferințe, doar oportunitățile diferă. Cu toate că s-ar părea că gusturile oamenilor în materie de muzică clasică diferă, diferențele observate în privința consumului se datorează (cel puțin așa se susține) faptului că oamenii au un „capital muzical destinat consumului" mai mare, obținând astfel o plăcere mai mare din muzica clasică decât alții. Firește, o ureche muzicală dezvoltată este mult mai asemănătoare unei abilități decât unei oportunități, însă este o abilitate care depinde de oportunitatea de a o dobândi. Dorința de a dobândi o anumită abilitate se presupune a fi aceeași pentru toți. Însă ultimul enunț scoate la iveală un punct slab al argumentului. În ziua de azi, accesul la muzica clasică difuzată la radio nu presupune, în principiu, niciun cost, chiar ținând cont de costurile de oportunitate (poți asculta în timp ce faci alte lucruri). Din moment ce nu toată lumea dorește să dobândească o ureche muzicală exersată, explicația trebuie să se găsească în înclinația subiectivă de a o dobândi, iar aceasta variază, printre altele, în funcție de impactul diferit al muzicii clasice asupra unor urechi *neexersate*.

Alte moduri în care oportunitățile sunt mai importante decât dorințele

Oportunitățile sunt mai importante dintr-un punct de vedere fundamental: sunt mai ușor de observat, nu doar de către cercetători, ci și de ceilalți indivizi dintr-o societate. Un principiu al strategiei militare spune că ar trebui să plănuiești în funcție de oportunitățile (verificabile) ale adversarului, și nu în funcție de intențiile sale (neverificabile).[1] Dacă avem motive să credem că inamicul *ar putea* avea intenții ostile, principiul ne poate face să presupunem ce e mai rău: inamicul ne va face rău dacă poate face acest lucru.[2] Situația se complică atunci când credința noastră în intențiile ostile ale dușmanului poate să își afle temeiul în percepția că *acesta* crede că noi avem mijloacele și, probabil, chiar intenția de a-i face rău. Într-o asemenea încurcătură, generată de subiectivitate, oportunitățile obiective pot apărea drept singurele care pot oferi o bază sigură pentru planificare.

---

[1] Supraestimarea puterii militare a sovieticilor sau a irakienilor ne aduce aminte, însă, că și oportunitățile pot fi foarte greu de verificat.

[2] Chiar dacă nu avem motive să credem că altă națiune nutrește intenții ostile în prezent, se poate face apel la planificarea de contingență, care are în vedere scenarii viitoare. Probabil că în arhivele Ministerului Apărării al Statelor Unite ale Americii se află adânc îngropat un plan de invadare a Canadei.

Un alt motiv pentru care oportunitățile pot părea mai fundamentale decât dorințele se leagă de posibilitatea de a influența comportamentul. De obicei, este mai ușor să schimbi circumstanțele oamenilor și oportunitățile acestora decât să schimbi ce au aceștia în minte.[1] Acesta este un argument de tip cost-beneficii privitor la eficiența de cost a unor politici alternative, nu un argument despre puterea explicativă relativă. Chiar dacă guvernul are o teorie bună, care îngăduie atât explicația, cât și predicția, se poate să nu îngăduie prea mult control din moment ce elementele asupra cărora poate acționa nu sunt cele mai semnificative din punct de vedere cauzal. Să presupunem că putem lega performanța economică slabă de oamenii de afaceri cu aversiune la risc și de sindicatele puternice. Guvernul poate fi convins că atitudinea mentală a managerilor este cea mai importantă cauză și, totuși, să nu fie în stare să facă ceva în această privință. Dimpotrivă, după cum s-a văzut sub administrațiile Reagan și Thatcher, sindicatele pot fi distruse prin acțiunea guvernului.

Să luăm în considerare un exemplu important: comportamentul suicidal. Pentru a comite un astfel de act nu este suficientă dorința de a te sinucide, ci trebuie găsite și mijloacele adecvate pentru a o face. Rata mare de sinucideri în rândul medicilor, de pildă, se poate datora în parte accesului facil la substanțe letale, care sunt metoda cel mai frecvent folosită pentru sinucidere în acest grup.[2] Cu toate că guvernul poate încerca să limiteze intențiile sinucigașe punând la dispoziție linii telefonice de asistență sau convingând mass-media să nu mai raporteze cazurile de suicid care pot declanșa sinucideri prin contaminare, cele mai eficiente rezultate se obțin îngreunând accesul la mijloacele de sinucidere.[3] Printre strategii se numără: construcția de bariere care fac mai dificile salturile de pe poduri sau clădiri înalte, controlul riguros al anumitor medicamente, restricții la vânzarea de pistoale, înlocuirea monoxidului de carbon, letal, cu gaz metan în cuptoarele de bucătărie și instalarea de convertori catalitici pentru reducerea emisiilor de monoxid de carbon la țevile de eșapament. În viitor, putem fi martorii cenzurării site-urilor de Internet care „oferă ajutor pentru sinucidere". Până și simpla înlocuire a sticluțelor de medicamente cu blistere cu folie a dus la reducerea numărului de sinucideri prin otrăvire cu paracetamol. Diminuarea numărului maxim de tablete pentru rețete individuale ar putea, de asemenea, să reducă șansele otrăvirilor severe. În Franța, dar nu și în Marea Britanie, conținutul fiecărei tablete de

---

1 În plus, după cum voi susține mai târziu, cel mai bun mod de a schimba ce au aceștia în minte este de a schimba circumstanțele.

2 În ciuda unei credințe răspândite, accesul polițiștilor la pistoale nu îi face mai susceptibili de a se sinucide în comparație cu alte categorii.

3 În anii 1970, rata sinuciderilor a scăzut în Marea Britanie, schimbarea fiind inițial atribuită liniilor telefonice de ajutor instituite de Centrele Samaritenilor, pentru a fi pusă ulterior pe seama înlocuirii, în consumul casnic, a gazului de cărbune cu gaz natural, mai puțin letal.

paracetamol a fost limitat prin lege la opt grame. Se crede că acesta este motivul pentru care hepatopatiile severe și morțile survenite în urma otrăvirii cu paracetamol se întâlnesc mai rar în Franța decât în Marea Britanie.

Firește, un individ hotărât va găsi o cale. Prin eliminarea unui mijloc obișnuit de a-și lua viața, scăderea ulterioară a ratei sinuciderilor poate fi temporară. Însă, în anumite cazuri măcar, efectul pare a fi fost, după cum era de așteptat, de lungă durată. Dacă impulsul de a te sinucide este trecător, și nu puternic înrădăcinat, acesta poate dispărea până la momentul în care un mijloc adecvat pentru sinucidere devine disponibil.[1] Prin urmare, simpla *întârziere* (și nu blocare) a accesului la mijloace ar putea fi eficientă pentru combaterea sinuciderilor comise impulsiv. Impunerea unui timp de așteptare înainte de cumpărarea unui pistol ar putea reduce atât sinuciderile, cât și rata criminalității.[2]

Varietăți ale interacțiunilor dintre dorințe și oportunități

Un exemplu mai complex al interacțiunilor dintre dorințe și oportunități poate fi luat din analiza pe care o face Madison facțiunilor în *The Federalist* (*Federalistul*). Acesta susține că, în regimul care ar ajunge să domine într-o democrație directă (care ar trebui să fie cu necesitate una mică) sau într-o republică reprezentativă mică, facțiunile ar avea atât motivele, cât și mijloacele pentru a produce daune. Dar, pe de altă parte, „în aproape fiecare caz, o pasiune comună sau un interes comun ar fi resimțite de majoritate". Pe de altă parte, numărul mic de cetățeni asigură oportunitatea pentru ca majoritatea să acționeze în mod discreționar împotriva minorității din motivul că aceștia se pot aduna în întâlniri mai ușor. Dimpotrivă, într-o republică mare este „mai puțin probabil ca o majoritate să aibă un motiv comun să încalce drepturile altor cetățeni sau, în cazul în care ar exista un astfel de motiv comun, ar fi mult mai dificil pentru toți aceia care îl cunosc să îi descopere forța și să acționeze la unison". În această transpunere de la nivel individual la nivel colectiv, argumentul bazat pe dorințe și oportunități ia o formă ușor diferită. Cu toate că, se gândea Madison, construcția instituțională nu este eficientă în schimbarea motivelor indivizilor, aceasta poate

---

[1] *Costul* sporit al depistării unor mijloace nu cred că ar putea preveni sinuciderea. S-ar putea să existe considerații legate de costuri și beneficii implicate în decizia de a-și lua viața, ca, de exemplu, compararea durerii pe care agentul o va provoca altora cu atenuarea propriei dureri prin actul sinuciderii, însă costul de a găsi un mijloc adecvat nu va aduce nicio schimbare în cazul unui individ hotărât.

[2] De fapt, cele mai multe state americane care impun o perioadă de așteptare înaintea achiziționării unui pistol o fac pentru a le da timp autorităților să verifice dacă acela care vrea să cumpere are cazier sau a suferit de vreo boală mintală, și nu pentru a-i da cumpărătorului posibilitatea să se mai gândească.

modifica șansele ca o majoritate de indivizi să împărtășească aceleași motive. Chiar dacă nu poate afecta oportunitățile membrilor unei majorități facționare (în cazul în care există așa ceva) de a acționa concertat, construcția instituțională le poate reduce *abilitatea* de a acționa, diminuând șansele ca aceștia să se cunoască unii pe alții. Înaintea epocii sondajelor de opinie probabil că au existat nenumărate ocazii în care o majoritate tăcută nu și-a dat seama de propria-i existență.

Există, însă, și un revers al medaliei în argumentul lui Madison: o republică mare nu va împiedica doar apariția unor majorități facționare, ci va împiedica și acțiunea concertată. În *Despre democrație în America* a lui Tocqueville regăsim un număr semnificativ de argumente bazate pe structura „nu doar". Putem lua ca exemplu discuția sa despre impactul sclaviei asupra proprietarilor de sclavi. Într-o primă instanță, sclavia este neprofitabilă în comparație cu munca salariată. „Muncitorul liber primește salariu, sclavul primește educație, mâncare, medicamente și haine; stăpânul își cheltuiește puțin câte puțin banii în sume mici pentru a-l sprijini pe sclav; pierderile de abia le observă. Salariul muncitorului este plătit tot odată și pare doar a-l îmbogăți pe cel care îl primește; dar, de fapt, cheltuiala cu sclavul e mult mai mare decât aceea cu omul liber, iar munca sclavului este mult mai puțin productivă."[1] Însă „influența sclaviei se extinde chiar mai departe, ea pătrunde în sufletul stăpânului și dă gusturilor și ideilor acestuia anumite înclinații." Pentru că munca este asociată cu sclavia, albii din Sud disprețuiesc „nu doar munca în sine, ci orice întreprindere în care succesul nu poate fi asigurat decât prin muncă". Acestora le lipsesc atât oportunitățile, cât și dorința de a se îmbogăți: „Sclavia (...) nu doar îi împiedică pe albi să își construiască averile, dar îi îndepărtează chiar și de la dorința de a vrea acest lucru." Dacă Tocqueville are dreptate, dezbaterea clasică asupra stagnării economice a societăților sclavagiste este greșită. Nu trebuie să ne mai întrebăm dacă lipsa dorinței de a investi sau lipsa oportunităților de a investi ne oferă explicația corectă întrucât ambele părți s-ar putea să aibă dreptate.

Argumentele lui Madison și Tocqueville au o structură comună: atât dorințele, cât și oportunitățile sunt modelate de aceeași a treia variabilă, pentru ca, la rândul lor, acestea să modeleze acțiunea (sau să o împiedice, în funcție de caz). În mod abstract, există patru posibilități (semnele plus și minus indică efecte cauzale pozitive, respectiv negative) (vezi Figura 9.2).

Cazul (A) este ilustrat de analiza lui Madison a democrațiilor directe sau a republicilor mici. Cazul (B) este exemplificat de argumentul în favoarea republicilor mari și de analiza lui Tocqueville a efectelor sclaviei asupra proprietarului de sclavi.

---

[1] Argumentul este opac într-un mod necaracteristic pentru autor. Un argument mai simplu este acela că, exceptând anumite ramuri ale agriculturii, sclavia este neprofitabilă pentru că nu creează niciun stimulent pentru ca sclavii să exceleze în munca depusă.

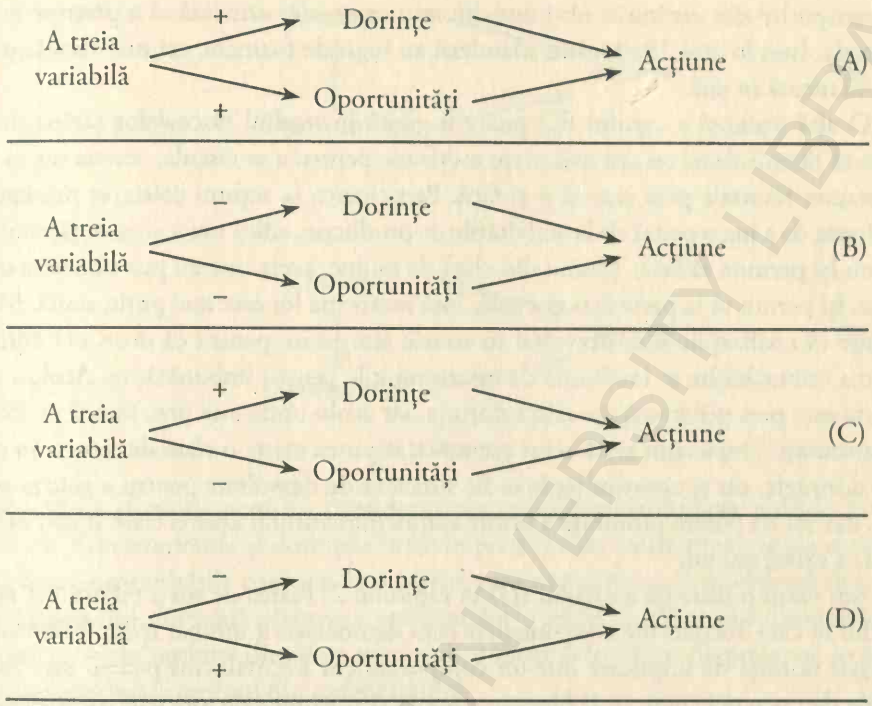

FIGURA 9.2

Cazul (C) se observă de multe ori când lipsa resurselor are efectul dual al creșterii stimulentelor pentru îmbunătățirea situației existente și al scăderii oportunităților în vederea acestei îmbunătățiri. Deși se spune că „nevoia este mama tuturor invențiilor", acest lucru este adevărat doar în măsura în care greutățile contribuie la creșterea motivației pentru a inova. Dar, din moment ce, de multe ori, inovarea presupune resurse (pe care le-am putea numi „tatăl invențiilor"), motivația singură poate să nu ducă nicăieri. De multe ori, inovarea cere investiții costisitoare, iar profitul poate fi nesigur sau poate veni foarte târziu – adică exact ce nu își permite o firmă pe cale să dea faliment. Firmele prospere își pot permite să inoveze, dar se poate să nu fie interesate să o facă. În cuvintele economistului John Hicks, „cel mai mare profit pe care îl aduce monopolul este o viață liniștită".

Similar, în timp ce dorința de a emigra este stimulată de sărăcia din țara natală, aceeași sărăcie poate fi responsabilă pentru blocarea accesului la mijloacele emigrării, date fiind costurile cu transportul. Până în prima jumătate a secolului al XIX-lea, emigranții spre Statele Unite își puteau folosi propriul corp drept gaj. Viitorii lor angajatori le plăteau drumul în schimbul unei perioade stipulate de serviciu. Astăzi, traficanții de persoane se pot baza pe frica de Serviciul pentru Imigrație și Naturalizare pentru a-i constrânge pe imigranții ilegali să nu își încalce promisiunea de a plăti costurile

BOSTON PUBLIC LIBRARY

cu transportul din veniturile obținute din munca pe care urmează să o presteze în țara-gazdă. Însă în anii 1840, când irlandezii au fugit de foamete, cei mai săraci au rămas să moară în țară.

O altă instanță a cazului (C) poate fi găsită în studiul răscoalelor țărănești: cu toate că țăranii săraci au cea mai mare motivație pentru a se răscula, aceștia nu au întotdeauna resursele prin care să o și facă. Participarea la acțiuni colective presupune abilitatea de a lua o pauză de la activitățile de producție, adică exact ceea ce țăranul sărac nu își permite să facă. Țăranii din clasa de mijloc, aceia care au pus câte ceva deoparte, își permit să ia parte la o răscoală, însă motivația lor este mai puțin acută. Marx susține că civilizațiile s-au dezvoltat în zonele temperate pentru că doar aici dorința pentru îmbunătățire se întâlnește cu oportunitățile pentru îmbunătățire. Acolo unde natura este prea generoasă nu există dorința, iar acolo unde este prea săracă nu există oportunități. După cum arată acest exemplu, ar putea exista o clasă de resurse în care atât dorințele, cât și oportunitățile să fie suficient de dezvoltate pentru a genera acțiune, dar nu ne putem pronunța a priori asupra dimensiunii acestei clase și nici măcar dacă ea există sau nu.

Am văzut o instanță a cazului (D) în capitolul 2. Partea de sus a Figurii 2.1 arată modul în care Tocqueville a argumentat cum democrația a inhibat (prin intermediul religiei) dorința de implicare într-un comportament necuviincios pentru care instituțiile democratice, cum ar fi libertatea presei și libertatea de asociere, au furnizat o oportunitate. O altă observație cunoscută a lui Tocqueville se bazează pe conjuncția lui (C) cu (D) și cu vârsta luată ca a treia variabilă: „În America, cei mai mulți oameni bogați încep prin a fi săraci; cei mai mulți oameni care duc o viață liniștită au muncit intens în tinerețea lor; prin urmare, la vârsta la care poți avea gustul studiului, aceștia nu au avut timpul necesar, iar când ajung să aibă la dispoziție timpul, le-a dispărut gustul."[1]

Dorințele și oportunitățile se pot și influența reciproc în mod direct: putem lua în considerare cazul (E) din Figura 9.3. În capitolul 2 am anticipat anumite moduri în care oportunitățile pot afecta dorințele: oamenii pot ajunge să dorească cel mai mult acel lucru pe care îl pot obține și pot obține sau prefera ceea ce au, în locul lucrului pe care nu îl au. Îl putem cita din nou pe Tocqueville cu privire la sclavie: „Este oare o binecuvântare de la Dumnezeu sau un blestem acea dispoziție a sufletului care le dă oamenilor un soi de gust depravat pentru cauza propriilor lor chinuri?" Acest mecanism

[1] O nepotrivire diferită este exprimată în proverbul francez „Si jeunesse savait, si vieillesse pouvait" (adică: „Dacă tinerii ar ști, dacă bătrânii ar putea"), care sugerează că două cerințe pentru acțiune, cunoașterea a ceea ce trebuie să faci și energia pentru a face, nu apar niciodată simultan. Un proverb norvegian pune problema ușor diferit: „Vârsta a doua este aceea în care ești suficient de bătrân pentru a ști că nu e bine să faci un lucru și suficient de tânăr ca totuși să îl faci."

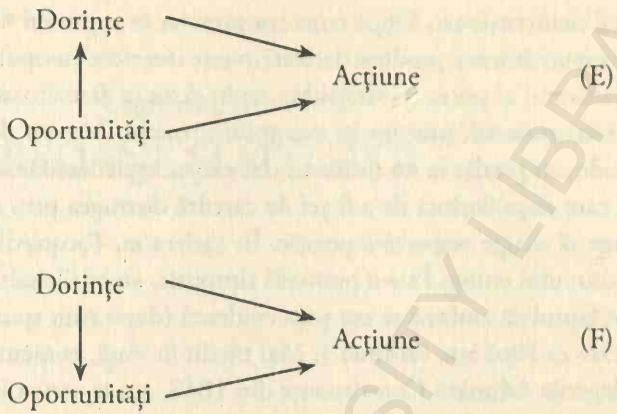

FIGURA 9.3

sugerează un alt motiv pentru a considera oportunitățile mai fundamentale decât preferințele. Oportunitățile și dorințele luate împreună sunt cauze proxime ale acțiunii, însă doar oportunitățile contează la o privire mai atentă întrucât modelează dorințele. Mecanismul „formării adaptive a preferințelor" (un fel de reducere a disonanței) se asigură ca nicio opțiune din afara setului de oportunități să nu fie preferată în locul celei mai preferate opțiuni din cadrul setului.

Ne putem întreba dacă nu cumva acest mecanism contează pentru comportament din moment ce, prin definiție, opțiunile care nu se află în setul de oportunități nu vor fi alese.[1] Să presupunem că agentul ordonează inițial opțiunile în ordinea A, B, C, D, iar apoi află că A nu este disponibil. Prin formarea adaptivă a preferințelor, persoana clasifică acum opțiunile în ordinea B, A, C, D. Ea va alege B, ca și cum preferințele sale ar fi rămas aceleași. Să presupunem, însă, că noua ordonare este C, B, A, D, determinându-se astfel o alegere pentru C. Acest lucru s-ar putea întâmpla printr-un proces de „supraadaptare" la numărul limitat de oportunități. Tocqueville susține că aceasta este ocaracteristică proprie francezului: „Acesta depășește cu mult spiritul obligației odată ce se vede obligat." Mai probabil, avem de a face cu o tendință generală observată în multe societăți ierarhice. De asemenea, noua ordonare a preferințelor poate fi B, C, D, A. Dacă femeile frumoase îmi resping avansurile mă pot consola cu gândul că, în virtutea narcisismului lor, acestea sunt cel mai puțin dezirabile partenere.

În cele din urmă, putem lua în considerare cazul (F), în care setul de oportunități este modelat de dorințele agentului. Acest lucru se poate întâmpla fie prin alegere intenționată, pe care o voi lua în considerare în capitolul 13, fie printr-un mecanism

[1] Adaptarea poate conta pentru *bunăstarea* agentului prin aceea că asigură liniștea sufletească. Acest efect este ambiguu, însă, pentru că auto-otrăvirea minții, asociată cu multe cazuri de „struguri acri" sau „lămâi dulci" (precum spectacolele de pe Broadway despre care am vorbit în capitolul 1), poate, într-o perspectivă mai largă, să scadă din bunăstare.

cauzal neintenționat. După cum am punctat în capitolul 4, dorința pentru stări care sunt esențialmente produse derivate poate interfera cu oportunitatea de a le produce. Mecanismul ar putea fi intrapsihic, ca în dorința dăunătoare sieși de a vrea să adormi, sau interpresonal, precum în exemplul următor. Într-un departament de universitate unde am predat la un moment dat exista regula tacită că oricare membru al facultății care afișa dorința de a fi șef de catedră distrugea prin aceasta oportunitatea de a ajunge să ocupe respectiva poziție. În cariera sa, Tocqueville a ilustrat ambele fațete ale ultimului enunț. Într-o perioadă timpurie, ambițiile sale politice au fost zădărnicite de faptul că ambiția sa era prea evidentă (după cum spune o vorbă, „Cine intră în conclav ca Papă iese cardinal"). Mai târziu în viață, comentând asupra succesului său în alegerile Adunării Constituante din 1848, acesta scrie că „nimic nu aduce succesul mai mult decât să nu ți-l dorești prea puternic".

◈◈◈

Notă bibliografică

Ideea „dorințelor irezistibile" este demolată într-un mod convingător de G. Watson în „Disordered appetites: Addiction, compulsion, and dependence", apărută în J. Elster (ed.), *Addiction: Entries and Exits* (New York, Russell Sage, 1999). Teza conform căreia indivizii au aceleași preferințe și diferă doar din punctul de vedere al oportunităților cu care se confruntă este asociată considerabil cu G. Stigler și G. Becker, „De gustibus non est disputandum", *American Economic Review* 67, (1977), pp. 76-90. Ideea preferințelor adaptive și, mai ales, aceea a supraadaptării la constrângeri i se datorează lui P. Veyne, *Le pain et le cirque* (Paris, Seuil, 1976). Declinul în rata (generală) a sinuciderilor care a urmat trecerii de la gazul de cărbune la gazul natural în Marea Britanie este documentată în N. Kreiman, „The coal gas story", *British Journal of Social and Preventive Medicine*, 30 (1976), pp. 86-93. Măsura în care accesul redus la mijloacele de sinucidere determină o utilizare mai mare a altor mijloace este discutată în C. Cantor și P. Baume, „Access to methods of suicide: What impact?", *Australian and New Zeeland Journal of Psychiatry*, 32 (1998), pp. 8-14. Utilizarea de către Madison a distincției oportunități-dorințe este analizată de M. White în *Philosophy, The Federalist, and the Constitution* (New York, Oxford University Press, 1987). Am discutat modul în care Tocqueville folosește această distincție în capitolul 4 din *Political Psychology* (Cambridge University Press, 1993).

Capitolul 10

PERSOANE ȘI SITUAȚII

Ruşinea şi vina sau disprețul și mânia diferă prin aceea că prima emoție a fiecărei perechi vizează *caracterul* persoanei, iar a doua privește o anumită *acțiune* realizată de către persoană (capitolul 8). Similar, aroganța se bazează pe credința superiorității în raport cu o altă persoană, iar mândria pe credința înfăptuirii unei fapte ieșite din comun. Însă când blamăm sau lăudăm o acțiune, oare nu o facem pentru că am crede că acțiunea reflectă caracterul agentului? Cărui alt factor i-am putea atribui toate acestea?

Unde greșește psihologia populară

Cartea de față nu este despre laudă sau blam, ci despre *explicația* comportamentului. Într-un astfel de context, se pune problema puterii caracterului de a explica acțiunea. Se presupune adesea că personalitatea oamenilor are anumite caracteristici (introvertit, timid etc.), precum și virtuți (onestitate, curaj etc.) sau vicii (cele șapte păcate capitale etc.). În psihologia populară se presupune că aceste trăsături sunt stabile în timp și în funcție de situație. Proverbe din toate limbile stau mărturie pentru această presupunere. „Cine spune o minciună va spune o sută." „Cel ce minte e și hoț." „Cine fură azi un ou mâine va fura un bou." „Cine știe să păstreze un secret mic știe să păstreze și unul mare." „Cine e prins cu mâța-n sac o dată va inspira mereu neîncredere." Dacă psihologia populară este corectă, predicția și explicația comportamentului ar trebui să fie simple. O singură acțiune va scoate la iveală trăsătura sau dispoziția subiacentă și ne va permite să prezicem comportamentul într-un număr indefinit de alte ocazii când se poate manifesta respectiva dispoziție. Procedura nu este tautologică, așa cum s-ar întâmpla dacă am lua ca dovadă de lipsă de onestitate trișarea la un examen, folosind apoi trăsătura de caracter a lipsei de onestitate pentru a explica trișarea. În schimb, o asemenea procedură presupune folosirea trișării la examen ca dovadă a unei trăsături de caracter (lipsa de onestitate), care poate, de asemenea, să determine persoana respectivă să nu fie fidelă partenerului de viață. Dacă acceptăm versiunea mai extremă a teoriei populare conform căreia toate virtuțile merg împreună, trișarea poate fi folosită pentru a prezice lașitatea în război sau băutul în exces.

BOSTON UNIVERSITY LIBRARY

Oamenii trag adesea concluzii tranșante doar din interpretarea comportamentului privat auster al altora. Este de notorietate că un membru al Academiei Franceze a votat pentru de Gaulle pentru demnitatea acestuia în viața privată, pornind de la premisa tacită că un politician capabil să își înșele soția este în stare să își trădeze și țara. În Vietnam, liderii comuniști au reușit să câștige „mințile și inimile populației" datorită stilului personal incoruptibil, în contrast cu acela, mai puțin ascetic, al liderilor din alte grupări politice. În rândurile Mafiei, aventurile extraconjugale sunt un semn de comportament necuvenit și de caracter slab.

Într-o oarecare măsură, psihologia populară devine realitate tocmai pentru că este așteptată. Dacă oamenii cred că alții le vor prezice comportamentul într-o situație de tip A în baza comportamentului lor într-o situație de tip B, atunci vor acționa în situația B având în minte situația A. Dacă credința existenței unei legături între moralitatea publică și cea privată este larg răspândită (iar acest lucru este cunoscut), atunci aceasta creează un stimulent pentru ca politicienii să se comporte onest în viața privată, presupunând că orice formă de comportament deplasat va ajunge la urechile electoratului. Ori să presupunem că este foarte răspândită credința conform căreia oamenii au aceeași rată a deprecierii timpului în toate situațiile. Dacă acestora le pasă foarte puțin de viitorul lor pentru a avea grijă de propriile corpuri, atunci, conform psihologiei populare, ar trebui să fie mai probabil ca aceștia să își încalce o promisiune de a realiza un câștig mare pe termen scurt. Prin urmare, pentru a putea să faci promisiuni credibile privitoare la o cooperare profitabilă pe termen lung trebuie să-ți mențiî silueta și să ai grijă de sănătatea ta.

Însă, pe termen lung, psihologia populară este *în mod demonstrabil falsă*.[1] Dacă vom putea elimina efectele psihologiei populare în sine, astfel încât să nu existe niciun stimulent pentru a satisface așteptările cerute de congruența dintre situații, vom vedea că nu rămânem cu un acord prea puternic. Părinții care observă comportamentul copiilor lor doar acasă sunt surprinși adesea să afle că aceștia sunt mult mai manierați la școală sau în vizită la colegii de clasă. Mai mult, intervențiile pentru îmbunătățirea comportamentului de acasă nu duc la o îmbunătățire a aceluia de la școală prin raportare la grupurile de control asupra cărora nu se exercită nicio intervenție.[2] În experimente de laborator, majorității participanților (cam două treimi dintre subiecți)

---

[1] Nu doar psihologia populară; și cercetătorii din științele sociale care susțin că agenții semnalizează prin comportament dacă fac parte din tipologia „celor buni" sau a „celor răi" supraestimează, de asemenea, relația dintre act și gând.

[2] Există un amendament la aceste descoperiri. Copiii părinților care s-au conformat sfaturilor date de intervenționiști au avut rezultate mai bune la școală decât copiii părinților care nu s-au conformat, fapt pe care unii îl citează în sprijinul efectului de răspândire dintre casă spre școală. Însă descoperirea se poate datora doar moștenirii conformismului. Este mult mai

li se poate induce un comportament necruțător, până în punctul în care li se impune aplicarea de (ceea ce aceștia cred a fi) șocuri electrice severe (cam 450 de volți) unui complice al aceluia care conduce experimentul. Însă nu avem vreun motiv să credem că respectivul comportament se datorează unei trăsături de caracter subiacente, fie aceasta sadismul, cruzimea sau indiferența la suferința celorlalți; de fapt, mulți dintre subiecții care s-au comportat astfel erau supărați pe ceea ce făceau și mâhniți. Copiii sunt mult mai dispuși să aștepte o recompensă venită mai târziu când recompensa respectivă și una de mai mici dimensiuni, pe care ar putea-o obține imediat, le sunt ascunse. Oricine din lumea universitară cunoaște confrați dedicați muncii lor, dar mai puțin conștiincioși când vine vorba de predat sau de ducerea la bun sfârșit a datoriilor administrative. A fi vorbăreț în timpul orei de prânz pare a fi o trăsătură slab corelată cu a fi vorbăreț în alte ocazii. O persoană poate amâna la nesfârșit să facă curățenie acasă, dar niciodată la locul de muncă.[1]

În eseul „Despre nestatornicia faptelor noastre", Montaigne pune în opoziție comportamentul lui Cato cel Tânăr cu acela al unor oameni obișnuiți, precum el însuși: „Dacă ai atins una dintre corzile lui Cato, atunci le-ai atins pe toate; există în acesta o armonie de sunete în perfect acord, pe care nu o poate nega nimeni. În ceea ce ne privește, dimpotrivă, toate acțiunile noastre trebuie judecate în mod independent: cel mai sigur mod, în opinia mea, ar fi să le punem în legătură cu circumstanțele lor imediate, fără să ne uităm mai departe și fără să tragem concluzii definitive." Acesta mai punctează: „Dacă un om nu poate suporta calomnia, dar își suportă sărăcia cu dârzenie, dacă nu poate să suporte bisturiul chirurgului, dar este neînfricat în fața săbiilor adversarului, atunci nu omul merită laude, ci fapta."

Voi da câteva exemple din lumea artei și a artiștilor. Proust scrie că „te-ai fi putut gândi" că tinerii bărbați din *Timpul regăsit* care fuseseră plătiți pentru a le face rău clienților bordelului lui Jupien trebuie să fie „răi într-un mod fundamental, însă nu doar că aceștia erau soldați minunați în timp de război, adevărați «eroi», ci fuseseră generoși și cumsecade în numeroase ocazii din viața publică". Comentând asupra contradicției evidente dintre „disimularea excelentă a unei invitații la Palatul Buckingham" și fala că soția unui funcționar de rang inferior a vizitat-o pe doamna Swann, acesta scrie că

---

probabil ca părinții care urmăresc cu conștiinciozitate instrucțiunile unui intervenționist să aibă copii care, la rândul lor, să urmeze conștiincios sfaturile date de un profesor.

[1] Într-o scrisoare către eticianul Randy Cohen (*New York Times Magazine*, 15 ianuarie 2006), o persoană din lumea universitară a pus întrebarea dacă cererea unui coleg care nu este membru al catedrei de a i se acorda reduceri la care nu este îndreptățit la Clubul Facultății impune un vot împotriva includerii sale în catedră „pentru lipsa lui de onestitate și probabilitatea ca aceasta să se extindă și asupra cercetărilor sale". Cohen a răspuns negativ, motivând că „oamenii care se comportă rău în anumite situații se comportă deseori foarte cuviincios în altele".

motivul cel mai important era (iar lucrul este valabil pentru întreaga umanitate) că nici măcar virtuțile noastre nu sunt elemente care plutesc undeva liber în afara noastră și care ne sunt întotdeauna la îndemână; de fapt, acestea ajung să fie atât de puternic legate în mintea noastră de acele ocazii pentru a acționa în care avem impresia că ar trebui să le manifestăm, încât, dacă ni se cere să ne angajăm într-o activitate diferită, aceasta ne poate lua prin surprindere, astfel că nu ne gândim vreun moment că și acest caz ar trebui să aducă după sine folosirea acelorași virtuți.

Muzicianul de jazz Charlie Parker a fost caracterizat de un doctor care îl cunoștea drept „un om care trăiește de la o clipă la alta. Un om care trăiește pentru principiul plăcerii, pentru muzică, mâncare, sex, droguri, senzații tari, personalitatea sa [*sic*] fiind blocată la un nivel infantil." Un alt muzician de jazz, Django Reinhardt, manifesta chiar o atitudine și mai extremă de orientare către prezent, neeconomisind niciodată nimic din câștigurile sale substanțiale, ci cheltuindu-le pe capricii și mașini scumpe, pe care le distrugea la scurt timp după aceea. Din mai multe puncte de vedere, acesta încarna stereotipul „țiganului". Și totuși, nu ajungi un muzician de calibrul lui Parker sau Reinhardt dacă trăiești clipa *din toate punctele de vedere*. Perfecționarea presupune ani de dedicație și de concentrare. În cazul lui Reinhardt, o astfel de atitudine a fost adusă la suprafață când și-a accidentat mâna stângă foarte grav într-un incendiu, fiind astfel constrâns să facă cu două degete ceea ce alții fac cu patru. Dacă acești doi muzicieni ar fi fost firi impulsive și delăsătoare în toate privințele – dacă „personalitatea" ar fi fost „infantilă" într-un mod consecvent –, atunci nu ar fi reușit niciodată să ajungă artiști desăvârșiți.

După 1945, romancierul norvegian Knut Hamsun, care a colaborat cu naziștii în război, a fost supus unor controale psihiatrice pentru a se stabili dacă este apt din punct de vedere mintal pentru a fi judecat (avea 86 de ani la momentul respectiv). Când profesorul de psihiatrie i-a cerut să-și descrie „trăsăturile fundamentale de caracter", replica romancierului a fost următoarea:

Așa-zisa perioadă naturalistă – Zola și epoca sa – au scris despre persoane cu trăsături de caracter. Nuanțele psihologice nu le erau de trebuință. Oamenii aveau o singură capacitate dominantă care le guverna acțiunile. Dostoievski și alții ne-au învățat pe toți ceva diferit despre oameni. De la bun început nu cred că există o singură persoană în oricare dintre scrierile mele care să aibă o trăsătură unică dominantă. Toți sunt lipsiți de așa-zisul caracter – sunt divizați și fragmentați, nu sunt buni sau răi, ci ambele. Sunt nuanțați și versatili atât în ceea ce gândesc, cât și în acțiunile lor. Fără îndoială că și eu sunt la fel. Este foarte posibil să fiu agresiv și să am și celelalte trăsături pe care le sugerează profesorul – sunt vulnerabil, suspicios, egoist, generos, gelos, cinstit, gândesc logic, sunt sensibil, am o

natură glacială. Toate acestea ar fi trăsături umane, dar nu-i pot acorda vreuneia o prevalență în propriul meu caracter.

Voi urmări în detaliu chestiunea caracterului sau a „lipsei de caracter" în capitolul 14, unde mă voi raporta la opere literare. Aici vreau doar să fac observația că Hamsun nu se referă la posibilitatea că ar putea fi, de pildă, generos în mod consecvent într-un anumit tip de situație și egoist în mod consecvent în altă situație. Mă voi apleca acum asupra acestei chestiuni.

Puterea situației

„A fi risipitor cu banii" și „a se dedica cu totul creației muzicale", „a fi vorbăreț în timpul orei de prânz" sau „a fi conștiincios în cercetare" constituie, desigur, trăsături de caracter. Sunt, însă, *trăsături locale* sau specifice situației, nu trăsături de personalitate globale, care să se manifeste indiferent de situația dată. Contrar psihologiei populare, studiile sistematice au descoperit niveluri foarte scăzute ale consecvenței de la caz la caz pentru trăsăturile de caracter. Cu toate că există corelații, acestea sunt în mod obișnuit atât de mici, încât nu pot fi detectate „cu ochiul liber". Psihopații pot manifesta un comportament nepăsător indiferent de situație[1], iar Cato cel Tânăr se poate să fi fost eroic în mod consecvent, dar, pentru cei mai mulți dintre indivizii aflați între aceste două extreme, nu trebuie să ne așteptăm la o astfel de consecvență. Ideea mai extremă a psihologiei populare, conform căreia toate virtuțile merg împreună, nu a fost testată cu atenție niciodată, poate pentru că pare atât de evident implauzibilă. Însă această idee poate avea încă un impact la nivel mental, lucru indicat de încrederea pe care o manifestăm față de abilitățile medicale ale doctorilor care se comportă frumos cu pacienții. În Antichitatea clasică era des întâlnită ideea că excelența într-un domeniu este un element predictiv infailibil sau un „index" al excelenței în alte domenii. Psihologii vorbesc în aceste cazuri despre un „efect de halo".

Prin urmare, de foarte multe ori explicația comportamentului poate fi găsită în *situație*, și nu în *persoană*. Să luăm în considerare, de pildă, faptul că au existat cetățeni germani care au contribuit la salvarea evreilor de regimul nazist. Într-o teorie „caracterologică", ar trebui să presupunem că salvatorii aveau un tip de personalitate altruist, care le lipsea acelora care nu au comis astfel de acțiuni. Însă, se pare că factorul cu cea mai mare putere explicativă este elementul „situațional", dat de faptul că *li s-a cerut* să salveze pe cineva. Legătura cauzală poate să apară în două moduri. Pe de o parte, au putut obține *informațiile* necesare pentru a se comporta ca niște salvatori doar pentru

[1] Din moment ce un om inteligent și egoist căruia îi pasă de viitor va avea adesea un interes în a *mima* preocuparea pentru alții (capitolul 5), explicația ultimă a comportamentului psihopaților poate sta în deprecierea excesivă a viitorului.

că au fost rugați. Pe de altă parte, situația de a fi rugați față în față să facă acest gest poate declanșa acceptare pentru a evita *rușinea* pe care ar simți-o altminteri.[1] Prima explicație presupune altruismul, dar neagă că acesta este suficient pentru a da o explicație comportamentului. Al doilea neagă altruismul și înlocuiește normele sociale cu normele morale. În oricare dintre cele două abordări, ceea ce face diferența între salvatori și ceilalți este situația în care aceștia se găsesc, și nu personalitatea lor.

Cazul lui Kitty Genovese este un alt exemplu din viața reală referitor la puterea situației. Este implauzibil să stipulăm doar pe baza inacțiunii lor că toți martorii la crimă au avut inima de piatră și au fost indiferenți în fața suferinței umane. Mai degrabă, mulți dintre ei au crezut probabil că altcineva a dat telefon la poliție sau că, din moment ce nimeni nu face nimic, situația nu este chiar atât de gravă pe cât pare („probabil că e doar o ceartă în familie") ori că inacțiunea celorlalți sugerează că intervenția directă poate implica riscuri.[2] Aceste raționamente devin mai plauzibile pe măsură ce crește numărul de martori. Astfel, în cadrul unui experiment, subiecții îl aud prin interfon pe complicele aceluia care face experimentul mimând o criză de epilepsie. Când subiecții cred că sunt singurii martori care aud, 85% dintre ei sar în ajutor; când cred că mai există încă un ascultător, 62% intervin; când cred că mai sunt patru care ascultă, 31% intervin. În alt experiment, 70% dintre martori intervin când sunt singuri, dar numai 7% intervin dacă stau lângă un complice impasibil. În cazul a doi subiecți naivi, victima a primit ajutor în 40% dintre cazuri. Astfel, nu doar șansa ca *oricare* dintre martori să intervină scade când sunt mai mulți, dar și șansa ca *un anumit* martor să intervină scade odată cu numărul de martori.[3] Cu alte cuvinte, diluarea responsabilității de a interveni, cauzată de prezența altora, se manifestă atât de rapid,

---

1 Similar, succesul teledonurilor în încurajarea oamenilor să dea bani nu stă în faptul că acestea stimulează altruismul, ci în faptul că sunt însoțite de o bătaie în ușă a cuiva care face o cerere față în față. Într-un astfel de caz, explicația care recurge la ideea de informație nu este adecvată.

2 Oamenilor cărora le-a fost frică să intervină fizic pentru a proteja victima de atacator puteau, totuși, să dea telefon la poliție. Însă, în acel moment, poliția nu accepta apeluri anonime, prin urmare, martorilor se poate să le fi fost frică să nu intre în necaz. În alte situații de acest fel, opțiunea de a suna la poliție poate să nu fie disponibilă.

3 Cel puțin aceasta este tendința generală în nenumăratele experimente de acest tip care au fost desfășurate. În cel tocmai menționat (criza de epilepsie auzită prin interfon) se dovedește că, dacă presupunem că ceilalți martori sunt subiecți reali, naivi care primesc aceeași informație (și nu simpli complici sau ficțiunile create pentru experiment), șansa ca măcar unul dintre ei să intervină este relativ constantă, adică în jur de 85%. În cazul celor cinci subiecți (subiectul central și cei patru ascultători), șansa ca oricare dintre ei să se abțină de la a interveni este 0,69. Șansa ca toți să se abțină este $(0,65)^5$ sau 0,156; rezultă astfel o probabilitate de 0,844 ca măcar unul dintre ei să intervină.

încât nu poate fi contracarată de numărul mai mare de oameni care pot interveni (potențial).

În alt experiment, unor studenți la teologie li s-a spus să se pregătească pentru o scurtă predică într-o clădire învecinată. Jumătate dintre ei au primit instrucțiuni să-și conceapă predica în jurul parabolei bunului samaritean (!), iar ceilalți să aleagă un subiect neutru. Unui grup i s-a spus să se grăbească întrucât oamenii din clădirea vecină îi așteaptă, pe când celorlalți li s-a spus că au timp suficient la dispoziție. În drum spre cealaltă clădire, subiecții întâlnesc un om căzut, rezemat de o ușă, având probleme, după toate aparențele. Doar 10% dintre studenții cărora li s-a spus că vor întârzia și-au oferit ajutorul; din celălalt grup, 63%. Grupul căruia i s-a spus să pregătească o predică pe tema bunului samaritean nu dădea semne că ar fi dispus să se comporte asemenea samariteanului. Iar comportamentul studenților nu s-a dovedit a fi corelat nici cu răspunsurile la un chestionar a cărui intenție era aceea de a măsura dacă interesul lor față de religie se datorează unei dorințe pentru mântuirea proprie sau unei dorințe de a-i ajuta pe ceilalți.[1] Factorul situațional – a fi grăbit sau nu – are o putere explicativă mult mai mare decât factorul dispozițional.

Nu ar fi corect să includem această analiză în aceea din capitolul trecut spunând că studenții „grăbiți" s-au comportat în modul în care au făcut-o din cauza unei *constrângeri de timp*. Constrângerea lor nu era una obiectivă sau „tare" și, de fapt, 10% dintre studenți s-au oprit să ofere ajutor. Mai degrabă, situația a modelat comportamentul, afectând dimensiunea *dorințelor* competitoare. Cererea față în față sporește tăria motivelor care țin cont de celălalt, pe când faptul că li s-a spus să se grăbească o diminuează. Faptul de a fi în măsură *să vezi* recompensa disponibilă iminent o face mai atractivă în comparație cu una care devine disponibilă mai târziu, la fel cum priveliștea unui cerșetor pe stradă poate declanșa o formă de generozitate pe care cunoașterea abstractă a sărăciei nu o poate declanșa. Situațiile de tipul Kitty Genovese schimbă atât costurile percepute, cât și beneficiile estimate ale ajutorului. Dorința de a te conforma instrucțiunilor date de organizatorul impasibil al unui experiment care îți spune că „trebuie să continui" să administrezi șocuri electrice, evident dureroase și, posibil, chiar fatale, anulează dorința de a nu provoca durere fără rost.

Nu există niciun mecanism general sau obișnuit prin care o situație poate afecta comportamentul. Situațiile variază de la cele mai importante, cum ar fi cerința față în față de a salva evrei, până la cele mai banale evenimente, ca atunci când descoperirea unei monede în fanta unui telefon public stimulează sentimentele de generozitate și ne face dornici să ajutăm un străin (în realitate, un complice în experiment) să adune niște hârtii pe care le-a scăpat pe trotuar. Lecția importantă pe care trebuie să

[1] Ca și subiecții care au fost convinși să provoace șocuri electrice, mulți dintre aceia care s-au grăbit să treacă pe lângă un om în necaz erau vizibil tulburați de întâlnire.

o extragem din aceste observații, din viața reală și din laborator, este aceea că, de fapt, adesea comportamentul nu este mai stabil decât situațiile care îl modelează. O persoană poate fi foarte locvace în timpul orei de masă, când poate să se relaxeze alături de colegii pe care îi cunoaște, dar poate fi foarte retrasă față de străini. O persoană poate să dea bani cerșetorilor în mod constant, dar să nu se gândească la săraci în alte împrejurări. O persoană poate acorda ajutor în situații în care nimeni altcineva nu sare în ajutor, dar poate fi pasivă în prezența altor oameni care și-ar putea oferi ajutorul. Un bărbat poate fi agresiv în mod constant și îi poate face observații dure soției sale, dar poate fi foarte calm și generos față de alți oameni. Și soția sa poate manifesta același comportament dual. Agresiunea lui o declanșează pe a ei și invers.[1] Dacă aceștia își văd partenerul de viață interacționând arareori cu alți adulți, de exemplu, la locul de muncă, pot ajunge să creadă că acesta sau aceasta este intrinsec agresiv(ă), și nu doar agresiv(ă) în situația definită de prezența lor.

Apelul spontan la dispoziții

Putem detalia ultimul exemplu. Cei care se ocupă de terapia de cuplu încearcă adesea să îi facă pe partenerii care le cer ajutorul să nu mai folosească limbajul centrat pe caracter și să se concentreze pe acela despre acțiune. În loc să spună „Ești o persoană rea", enunț care lasă puțin loc pentru speranță sau schimbare, ei ar trebui să facă efortul să spună „Ai făcut un lucru rău". Ultimul enunț lasă deschisă posibilitatea ca acțiunea în cauză să fi fost declanșată de factori specific situaționali, precum o remarcă provocatoare a partenerului de viață. Un motiv (printre altele) pentru care terapeuții nu au prea mare succes în reîncadrarea conflictelor în acest mod este că oamenii acordă o mai mare atenție în mod spontan acelor explicații ale comportamentului care se bazează pe caracter, față de acelea care se bazează pe situații. Dacă aflăm că cineva a contribuit la o reclamă pentru „drepturile homosexualilor", avem tendința să credem că persoana este gay sau că are o gândire liberală, și nu că aceasta *a fost rugată* într-un mod greu de refuzat. În cazul unui interviu pentru un loc de muncă, avem tendința să explicăm ce face sau spune o persoană în lumina dispozițiilor pe care i le atribuim (prea încrezători în noi înșine), și nu naturii speciale a situației de interviu. Limba însăși reflectă părtinirea pe care o acordăm dispozițiilor. Adjectivele care se aplică acțiunilor („ostilă", „egoistă" sau „agresivă") pot fi, de asemenea, aplicate agentului, în timp ce există foarte puține caracterizări ale acțiunilor care pot fi aplicate și situațiilor („dificilă" este o excepție).

---

[1] Vorbind metaforic, putem spune că aceștia se află într-un „echilibru psihologic defectuos". Însă agresiunea nu trebuie să fie „cel mai bun răspuns" dat agresiunii (lucru sugerat de noțiunea de echilibru din teoria jocurilor), ci trebuie doar să fie inteligibilă din punct de vedere psihologic.

Psihologii numesc utilizarea improprie a explicațiilor dispoziționale *eroarea fundamentală de atribuire,* adică explicarea comportamentului indus de o situație prin faptul că este cauzat de trăsăturile de caracter persistente în timp ale agentului. Subiecții cărora li s-a cerut să prezică comportamentul studenților la teologie care întâlnesc un individ aflat în nevoie au crezut (în mod greșit) că oamenii a căror religie se bazează pe o dorință de a-i ajuta pe alții au o probabilitate mai mare de a se comporta precum bunul samaritean și (de asemenea, greșit) că nu are importanță dacă sunt mai mult sau mai puțin grăbiți. Alți subiecți au făcut o predicție prea tare cu privire la administrarea șocurilor electrice în absența unor factori situaționali specifici în experimentul original, dându-și astfel la iveală credința într-o explicație dispozițională. Când un instructor îi dă unui student sarcina de a scrie un eseu în care să argumenteze în favoarea lui Castro, alți studenți, cunoscând modul în care s-a cerut tema, interpretează respectivul eseu drept o manifestare a unei atitudini favorabile lui Castro. Când studenții au fost rugați să se ofere voluntari pentru sarcini a căror remunerație era fie mică, fie mare, iar înrolarea a fost mică, respectiv mare, observatorii, știind de existența unor diferențe de plată, au prezis, cu toate acestea, că este mai probabil ca *toți* voluntarii să se înroleze pentru o cauză neplătită, spre deosebire de non-voluntari. Cu alte cuvinte, observatorii au atribuit acțiunea voluntariatului unei dispoziții a voluntarului mai degrabă decât structurii de recompensă a situației.

În anumite societăți, oamenii par mult mai puțin dispuși decât aceia din alte societăți să comită eroarea fundamentală de atribuire. Experimentele indică faptul că, în comparație cu americanii, asiaticii acordă o mai mare importanță situației decât unor dispoziții personale în explicarea comportamentului. Diferența de atribuire poate fi pusă în evidență și prin exemple din viața reală. Astfel, în 1991, un student chinez la o facultate de fizică și-a împușcat coordonatorul, câțiva colegi, iar apoi s-a sinucis. În același an, un angajat al Poștei din Statele Unite care și-a pierdut locul de muncă și-a împușcat șeful, câțiva colegi și martori, iar apoi s-a sinucis. Ambele evenimente au fost pe larg mediatizate în ziarele de limbă engleză, precum și în cele de limbă chineză, primele explicând evenimentele preponderent prin termeni dispoziționali („tulburat", „furios", „labil psihic"), iar cele din urmă în temeni situaționali („acces ușor la armă", „tocmai își pierduse locul de muncă", „victimă a presiunii de a reuși"). Alte descoperiri confirmă existența unei diferențe. Totuși, acest lucru se poate datora faptului că factorii situaționali joacă un rol mai mare în generarea comportamentului asiaticilor. În loc să fie mai eficienți cu privire la învingerea caracterului părtinitor al dispozițiilor, aceștia s-ar putea să nu mai aibă ce să învingă. S-ar putea, de asemenea, ca ambii factori să funcționeze.

Depășirea erorii fundamentale de atribuire poate fi eliberatoare. Studenții din anul întâi cărora li se spune că majoritatea bobocilor au rezultate slabe, dar că notele lor se vor îmbunătăți pe parcurs, chiar stau mai bine la note în anii ulteriori, în comparație

cu aceia cărora nu li se dă această informație. Probabil că aceștia din urmă atribuie rezultatele slabe lipsei de abilități, și nu mediului nefamiliar și derutant pe care îl reprezintă viața de campus. Grupurile supuse opresiunii care se debarasează de esențialismul acelora care îi asupresc – ideea că femeile, persoanele de culoare sau evreii sunt intrinsec inferiori – ar putea mult mai lesne să se debaraseze de lanțurile opresiunii.

Este eroarea fundamentală de atribuire „fierbinte sau rece" o greșeală motivată sau ceva asemănător mai degrabă cu o iluzie optică? În măsura în care motivația intră în procesul atribuirii, nu există niciun motiv care să ne determine să punem un accent prea puternic pe dispoziții. Pe temeiuri care servesc doar propriului nostru interes, ar trebui să atribuim succesul nostru unor trăsături solide de caracter, iar eșecurile ar trebui să le punem pe seama unor circumstanțe nefericite.[1] Dacă ar fi să-i credem pe moraliștii francezi, ar trebui să atribuim succesele celorlalți norocului, iar eșecurile lor dispozițiilor.[2] Pe temeiuri cognitive, tendința de a favoriza persoana mai mult decât situația poate fi o instanțiere a unei tendințe mai generale, aceea de a acorda o atenție sporită prim-planului mai alert, și nu fundalului static. Rezultă că eroarea ar trebui să fie mai puțin obișnuită în culturile în care atenția pentru prim-plan și, respectiv, cea pentru fundal sunt ceva mai echilibrate. Astfel par să stea lucrurile în culturile asiatice.

Reabilitarea persoanei

Observațiile pe care le-am descris subminează ceea ce am putea numi „esențialismul crud" în studiul personalității. Pur și simplu, nu este adevărat că oamenii *sunt* în sine agresivi, nerăbdători, extrovertiți sau vorbăreți. În același timp, nu decurge din aceste descoperiri experimentale că situația este atotputernică în explicarea comportamentului. Trebuie, mai degrabă, să descompunem caracterul, pe care în mod obișnuit îl înțelegem ca un „tot", într-un set de tendințe de răspuns *contingente*. În loc să caracterizăm pe cineva drept altruist, mai bine l-am caracteriza prin expresia „ajută dacă este solicitat, dar nu solicită să ajute" sau prin expresia „ajută dacă nu este stresat, dar este neglijent dacă este stresat". Fiecare dintre aceste expresii ar putea caracteriza un aspect al persoanei, putând astfel să circumscrie o formă mai subtilă de esențialism. Cineva poate să își mustre partenerul de viață pentru că nu face curățenie prin casă („ești leneș") sau pentru că nu face curat decât atunci când i se cere („ești nepăsător"). În cel din urmă caz, partenerul de viață poate fi proactiv, și nu reactiv în alte privințe,

---

[1] Însă, uneori, putem fi motivați să atribuim eșecurile noastre caracterului. Un parior sau un alcoolicse poate auto-liniști spunându-și că „nu se poate abține" pentru a-și furniza o scuză să continue.

[2] Putem fi motivați să atribuim succesele celorlalți *defectelor* lor de caracter. Antisemitismul se sprijină pe mitul că evreii au succes pentru că au un caracter imoral, care îi face capabili să adopte orice mijloace pentru a reuși.

cum ar fi faptul de a avea grijă de sănătatea copiilor. Nu am avea de-a face cu trăsătura de caracter atribuibilă reactiv în orice situație.

Din această perspectivă, explicația comportamentului se bazează pe situația particulară însumată cu relația specific personală dintre situații și cu înclinația comportamentală. Cineva poate fi puternic agresiv cu indivizi asupra cărora are putere, dar excepțional de prietenos cu aceia care au putere asupra sa, în timp ce altă persoană poate manifesta trăsături complet opuse. Dacă observăm că ambele persoane se comportă prietenos, putem fi tentați să tragem concluzia că ambii *au* o dispoziție prietenoasă. Însă, în virtutea celor argumentate până acum, ar trebui să fie limpede că asemănările de comportament se pot datora diferențelor cu privire la situație și la elementele contingente de răspuns, care se anulează reciproc.

◈◈◈

Notă bibliografică

„Tendința de a supraestima unitatea personalității" a fost foarte clar enunțată de G. Ichheiser, „Misunderstandings in human relations: A study in false social perception", *American Journal of Sociology*, 55 (1949), anexă. Studiile mai recente mută accentul de pe „caracter" înțeles ca un tot, ca la W. Mischel, *Personality and Assessment* (New York, Wiley, 1968). Expunerea de față se bazează substanțial pe L. Ross și R. Nisbett, *The Person and the Situation* (Philadelphia, Temple University Press, 1991), și pe J. Doris, *Lack of Character* (Cambridge University Press, 2002). Referințele la liderii comuniști din Vietnam și la Mafie îi aparțin lui S. Popkin, *The Rational Peasant* (Berkeley, University of California Press, 1979) și, respectiv, lui D. Gambetta, „Trust's odd ways", în J. Elster *et. al.* (editori), *Understanding Choice: Explaining Behavior: Essays in Honour of Ole-Jørgen Skog* (Oslo, Academic Press, 2006). Referința la efectele intervenționismului este din J.R. Harris, *No Two Alike* (New York, Norton, 2006). Disponibilitatea de a administra șocuri electrice este descrisă în studiul clasic al lui S. Milgram, *Obedience to Authority* (New York, Harper, 1983). Informațiile despre cei doi muzicieni pot fi găsite în R. Russell, *Bird Lives: The High Life and Hard Times of Charlie (Yardbird) Parker* (New York, Charterhouse, 1973), și în M. Dregni, *Django: the Life and Music of a Gypsy Legend* (Oxford University Press, 2004). Declarația lui Hamsun este tradusă din G. Langfeldt și Ø. Ødegård, *Den rettspsykiatriske erklæringen om Knut Hamsun* (Oslo, Gyldendal, 1978), p. 82. Folosirea comportamentului ca un „index" în Antichitate este discutată în P. Veyne, *Le pain et le cirque* (Paris, Seuil, 1976), pp. 114, 773; cf. și P. Veyne, „Pourquoi veut-on qu'un prince ait des vertus privées?", *Social Science Information*, 37 (1998), pp. 407-415. Explicația „caracterologică" a disponibilității de a salva evrei este susținută de K. Monroe, M.C. Barton și

U. Klingemann, „Altruism and the theory of rational action: Rescueres of Jews in Nazi Europe", *Ethics*, 101 (1990), pp. 103-122. Explicația „situaționistă" este susținută de F. Varese și M. Yaish, „The importance of being asked: The rescue of Jews in Nazi Europe", *Rationality and Society*, 12 (2000), pp. 307-324. În A. Dixit și S. Skeath, *Games of Strategy*, ediția a doua (New York, Norton, 2004), pp. 414-418, se susține, din perspectiva teoriei jocurilor aplicată cazului Kitty Genovese, că este de așteptat ca șansele ca *oricine* să intervină să scadă când crește numărul de persoane care pot interveni. O poziție sceptică despre tendința de a infera dispoziții bazându-ne pe comportament se face remarcată în J.L. Hilton, S. Fein și D. Miller, „Suspicion and dispositional inference", *Personality and Social Psychology Bulletin*, 19 (1993), pp. 501-512. Contrastul dintre americani și asiatici este rezumat în R. Nisbett, *The Geography of Thought* (New York, Free Press, 2004). Ceea ce am numit aici „reabilitarea persoanei" este prezentată în W. Mischel, „Towards an integrative science of the person", *Annual Review of Psychology*, 55 (2004), pp. 1-22.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 11

ALEGEREA RAȚIONALĂ

Structura alegerii raționale

Teoriile alegerii raționale vor să explice comportamentul bazându-se pe simpla presupunere că agenții sunt raționali. Prin urmare, nu este nevoie să clasificăm elementele determinante ale comportamentului ca fiind ori subiective (dorințe), ori obiective (oportunități). Teoria alegerii raționale este subiectivă în întregime.

Structura explicației prin alegerea rațională este prezentată în Figura 11.1. O acțiune este rațională, conform acestei scheme, dacă satisface *trei cerințe de optimalitate*: acțiunea trebuie să fie optimă, date fiind credințele, credințele trebuie să fie susținute cât de bine posibil, date fiind probele, iar probele trebuie să fie rezultatul unei investiții optime în culegerea de informații. În Figura 11.1 săgețile au o dublă interpretare: în termeni de cauzalitate și în termeni de optimalitate. De pildă, acțiunea ar trebui să fie cauzată de dorințele și credințele care o fac rațională; nu este suficient să faci ceea ce trebuie din pură întâmplare. Similar, o credință nu este rațională dacă este rezultatul a două procese părtinitoare opuse, care se anulează reciproc.

Pentru a da un exemplu, atât fumătorii, cât și nefumătorii procesează informații cu privire la pericolele fumatului în moduri care îi fac să creadă că acestea sunt

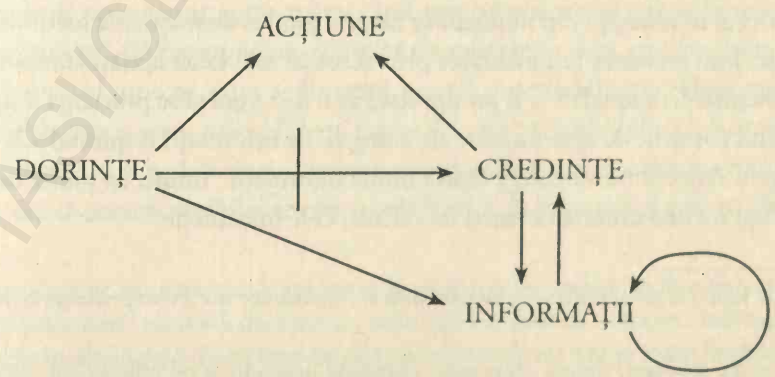

FIGURA 11.1

BCU IAȘI CENTRALA UNIVERSITY LIBRARY

mai mari decât în realitate. În același timp, fumătorii sunt supuși unei părtiniri care servește propriilor interese și îi face să deprecieze riscurile. Dacă rezultatul este acela că își formează aceeași opinie cu privire la fumat ca și un observator nepărtinitor[1], atunci acest lucru nu dovedește că sunt raționali. Într-una dintre cele mai influente discuții cu privire la raționalitate în științele sociale, Max Weber face greșeala de a infera „raționalitatea procesuală" din „optimalitatea rezultatului", susținând:

> Pentru scopurile unei analize științifice tipologice este convenabil să tratăm toate elementele de comportament iraționale și determinate de afecte ca factori de deviație de la tipul pur din punct de vedere conceptual al acțiunii raționale. De exemplu, un moment de panică la bursă poate fi analizat cel mai convenabil prin faptul de a încerca să determinăm care ar fi fost cursul acțiunii dacă aceasta nu ar fi fost afectată de afecte iraționale; este posibil, apoi, să introducem componentele iraționale ca fiind acelea care dau seama de deviațiile observate de la acest curs ipotetic al lucrurilor. Asemănător, în analiza unei campanii politice sau militare este convenabil să determinăm în primul rând care ar fi fost cursul rațional, date fiind scopurile participanților și cunoștințele adecvate în toate circumstanțele. *Doar în acest mod* este posibil să evaluăm semnificația cauzală a factorilor iraționali, ca dând seama de deviații de la acest tip.

Cu toate că Weber are dreptate să considere că deviațiile de la cursul rațional al acțiunii sunt o condiție suficientă pentru a avea de-a face cu iraționalitatea, acesta greșește afirmând (expresia pe care am pus-o în italice) că sunt și o condiție necesară. O greșeală asemănătoare este implicată în afirmația că reacțiile provocate de frica instinctivă sunt raționale, când tot ceea ce se poate spune despre acestea este că sunt *adaptive*. Când văd un obiect în drumul meu care ar putea fi un băț sau un șarpe, are sens să fug imediat, nu să culeg în continuare informații. Se pare că ființele umane sunt proiectate pentru a face acest lucru. Acest comportament nu este rațional în sensul strict al termenului din moment ce nu este produs de mașinăria luării de decizii raționale, însă mimează raționalitatea prin aceea că este exact același comportament pe care mașinăria rațională l-ar fi produs dacă ar fi fost pusă să se pronunțe asupra situației. Când costurile de oportunitate ale culegerii de informații (capitolul 12) sunt mari, un agent rațional nu va culege foarte multe informații. Totuși, de multe ori, tendința de a fugi nu este cauzată de astfel de calcule, ci le înlocuiește.[2]

---

[1] De fapt, cel de-al doilea tip de părtinire al fumătorilor nu îl compensează în întregime pe primul.

[2] Pentru șobolani, timpul scurs între răspunsul negândit și cel reflexiv este de aproximativ 10 milisecunde.

Preferințe și utilitate ordinală

Mai pe larg, prima cerință de optimalitate este că acțiunea trebuie să fie cel mai bun mijloc pentru a satisface dorințele agentului, date fiind credințele despre opțiunile disponibile și consecințele lor. Ce este „cel mai bine" este definit în termeni de „superioritate" sau preferințe: cel mai bun lucru este acela care nu admite altul mai bun, conform judecății agentului. Nu există nicio presupoziție ca dorințele să fie egoiste. Confundarea raționalității cu egoismul este o eroare grosolană, cu toate că este înlesnită de practica unor teoreticieni ai alegerii raționale. Dar nici nu trebuie să impunem cerința ca dorințele să fie stabile, nici măcar în sensul minimal al excluderii schimbărilor temporare de preferințe. Un agent aflat sub influența unei emoții sau a unor droguri preferă A lui B și se comportă rațional în alegerea lui A, chiar dacă, în alte circumstanțe, ar prefera B lui A. Un astfel de caz (vezi capitolul 6) apare atunci când greutatea atribuită de agent unor consecințe viitoare ale alegerii prezente este diminuată ca rezultat al unor astfel de influențe.

Pentru a pune în mișcare această analiză, noțiunea de „cel mai bun" trebuie să fie bine definită. Ne asigurăm de acest lucru punând două condiții.[1] În primul rând, preferințele trebui să fie tranzitive. Să presupunem că avem trei opțiuni: A, B și C. Dacă o persoană crede că A este cel puțin la fel de bun ca B, iar B este cel puțin la fel de bun precum C, atunci persoana ar trebui să creadă că A este cel puțin la fel de bun precum C. Dacă tranzitivitatea eșuează – de exemplu, dacă persoana preferă în mod strict pe A lui B, pe B lui C și pe C lui A –, poate că aceasta nu are „cea mai bună" opțiune. Mai mult, o altă persoană poate exploata această slăbiciune oferind agentului să își modifice o opțiune preferată mai puțin cu una preferată mai mult în schimbul unei sume de bani. Din moment ce preferințele funcționează ciclic, o astfel de operațiune poate fi repetată la nesfârșit, ruinând persoana respectivă printr-o serie de îmbunătățiri treptate.[2]

O situație de acest gen ar putea apărea când agentul ordonează opțiunile prin „numărarea de aspecte". Să presupunem că prefer un măr altuia dacă este mai bun în cel puțin două din trei aspecte, cum ar fi prețul, gustul și perisabilitatea. Dacă mărul A bate mărul B la preț și gust, mărul B bate mărul C la preț și perisabilitate, iar mărul C bate mărul A la gust și perisabilitate, atunci avem de-a face cu o încălcare a tranzitivității. Cu toate că această posibilitate este relativ lipsită de importanță pentru alegerea

---

[1] Ignor condiția de natură tehnică, anume că setul de opțiuni trebuie să fie compact și închis. Compactitatea este încălcată dacă pentru orice opțiune fezabilă există una mai bună, iar închiderea este încălcată dacă limita unui șir de opțiuni fezabile nu este ea însăși fezabilă.

[2] Pentru alt exemplu de „a te îmbunătăți până la moarte", vezi nota 1 din capitolul 18.

individuală care reflectă doar eșecul unei reguli empirice, vom vedea (capitolul 25) că este mult mai semnificativă pentru alegerea colectivă.

Suntem puși în fața unei probleme diferite în cazul în care indiferența eșuează în a fi tranzitivă. Pot fi indiferent între A și B și între B și C pentru că diferențele pentru fiecare pereche sunt prea mici pentru a fi sesizabile, dar prefer pe C lui A pentru că există o diferență detectabilă între ele. Există o opțiune care este „cea mai bună", respectiv C, dar este încă posibil să înrăutățim situația agentului făcându-i o serie de oferte – schimbând C cu B și B cu A –, astfel încât acesta nu are vreun motiv să refuze, prin urmare, poate foarte bine să accepte. Considerarea unui agent cu preferințe intranzitive drept irațional este justificată nu atât de lipsa „celei mai bune" opțiuni, ci de faptul că acesta poate accepta oferte care îi înrăutățesc starea.

Pentru a ne asigura că ideea de „cel mai bun lucru" are întotdeauna sens, trebuie să cerem ca preferințele să fie *complete*: pentru oricare două rezultate, agentul trebuie să fie în măsură să spună dacă preferă primul rezultat în locul celui de-al doilea, dacă preferă al doilea rezultat în locul primului sau dacă este indiferent între cele două rezultate. Dacă agentul nu este în măsură să dea oricare dintre aceste trei răspunsuri, este posibil ca acesta să nu fie în măsură să determine care este cea mai bună opțiune. Voi aborda lipsa caracterului complet spre sfârșitul capitolului. Pentru moment, vreau doar să fac observația că, spre deosebire de lipsa tranzitivității, lipsa caracterului complet nu este vreun soi de eșec. Să presupunem că vreau să îi dau unuia dintre copiii mei o înghețată, anume aceluia care se va bucura cel mai mult de ea. Pentru ca *eu* să am o preferință privitoare la cele două opțiuni ar trebui să fiu în măsură să compar nivelurile *lor* de satisfacție privitoare la preferințe în cazul în care li s-ar da înghețata respectivă. Însă, de cele mai multe ori, aceasta este o sarcină imposibilă. Însă faptul că nu putem duce la bun sfârșit această sarcină, în sensul că aș fi putut cumva să fac o alegere mai bună, nu constituie un eșec, ci reflectă un simplu fapt al vieții.

În multe cazuri, tranzitivitatea și caracterul complet al preferințelor sunt suficiente pentru a identifica acțiunea rațională. Însă adesea este convenabil să reprezentăm preferințele prin numere, numite de obicei *valori de utilitate*, pe care le atribuim opțiunilor. Pentru a ne asigura de această posibilitate trebuie să impunem o condiție suplimentară asupra preferințelor: *continuitatea*. Dacă fiecare opțiune dintr-un șir A1, A2, A3, ... este preferată lui B, iar șirul converge spre A, atunci A ar trebui să fie preferat lui B; dacă B este preferat în locul fiecărei opțiuni din șir, B ar trebui să fie preferat lui A. Un contraexemplu este furnizat de „preferințele lexicografice": un pachet constituit din două bunuri A și B în cantitățile (A1, B1) este preferat unui alt pachet (A2, B2) dacă și numai dacă fie A1 > A2, fie (A1 = A2 și B1 > B2). Într-o astfel de ordonare a preferințelor, pachetele (1,1, 1), (1,01, 1), (1,001, 1), ... sunt toate preferate lui (1, 2), care este preferat lui (1, 1). Putem spune că prima componentă a pachetului este incomparabil mai importantă decât a doua din moment ce, indiferent de cât de

mult am suplimenta cu bunul B, nu putem compensa pentru cea mai mică pierdere din bunul A.[1] Sau, mai simplu, nicio compensație nu este posibilă. Prin urmare, nu putem reprezenta aceste preferințe prin curbe de oportunitate. În timp ce preferințele lexicografice se aplică rar (dacă se aplică) bunurilor de consum obișnuite, acestea pot conta pentru alegeri politice. Un votant îl poate prefera pe candidatul A lui B dacă și numai dacă A are o atitudine puternică împotriva avortului sau dacă respectivii candidați au aceeași opinie împotriva avortului, dar A propune taxe mai mici decât B. Pentru astfel de votanți, „valoarea sacră" a vieții nu poate fi compensată de valoarea seculară a banilor.

Dacă preferințele agentului sunt complete, tranzitive și continue, putem să le reprezentăm printr-o funcție continuă de utilitate u care atribuie un număr $u(A)$ fiecărei opțiuni (A). În loc să spunem că un agent rațional alege cea mai bună opțiune fezabilă, putem spune că agentul își *maximizează utilitatea*. În această expresie, „utilitatea" este doar un mod convenabil de a denumi preferințele cu anumite proprietăți. Pentru a ne da seama de acest lucru, putem observa că singura cerință pentru ca o funcție u să reprezinte o ordonare de preferințe este aceea ca A să fie preferată lui B dacă și numai dacă $u(A) > u(B)$. Dacă u este întotdeauna mai mare ca zero, $v = u^2$ poate reprezenta, de asemenea, aceeași ordonare de preferințe, cu toate că v atribuie numere mai mari decât u, respectiv mai mici decât u în cazul în care $u < 1$. Numerele nu au nicio semnificație în valoare absolută, ci doar în termeni de magnitudine relativă sau *ordinală*. Prin urmare, ideea „maximizării utilității" nu implică faptul că agentul se angajează în a obține cât mai mult posibil din „ceva" de natură psihică. Se exclude, însă, tipul de ierarhie valorică de tipul preferințelor lexicografice. De fapt, cele din urmă nu pot fi reprezentate de o funcție de utilitate.

Utilitate cardinală și atitudini față de risc

De multe ori, agenții stau în fața unor opțiuni *riscante*, adică alegerile pe care le fac și ale căror probabilități sunt cunoscute au mai mult de un rezultat posibil. Intuitiv, s-ar părea că un agent rațional ar alege opțiunea cu cea mai mare utilitate așteptată, o idee care încorporează utilitatea fiecărui rezultat, precum și probabilitatea aparițiilor. Agentul ar trebui mai întâi, pentru fiecare opțiune, să pondereze utilitatea fiecărei consecințe prin probabilitatea sa și să adune toate utilitățile astfel ponderate, iar apoi să aleagă opțiunea a cărei sumă este cea mai mare.

Dar utilitatea ordinală *nu* ne îngăduie să explicităm această idee. Să presupunem că sunt date două opțiuni, A și B. A poate produce rezultatul O_1 sau O_2 cu probabilitățile ½ și ½, pe când B poate produce rezultatele O_3 sau O_4 cu probabilitățile ½ și ½.

---

[1] Noțiunea intuitivă de incompatibilitate poate fi exprimată, prin urmare, în două moduri distincte: ca preferințe incomplete sau ca preferințe discontinue.

Să presupunem acum că o funcție de utilitate u atribuie valorile 3, 4, 1 și, respectiv, 5 lui O_1, O_2, O_3 și, respectiv, O_4. „Utilitatea ordinală așteptată" a lui A este 3,5 și a lui B este 3. Dacă, însă, înlocuim funcția prin $v = u^2$, numerele sunt 12,5 și 13. Fiecare funcție reprezintă preferințele la fel de bine, însă fiecare identifică opțiuni diferite ca fiind „cele mai bune". Este evident că o astfel de abordare nu ne este utilă.

Putem obține rezultate mai bune, dar cu un cost la nivel conceptual. Varianta abordată de John von Neumann și Oskar Morgenstern arată că putem atribui valori de utilitate opțiunilor care nu au doar o simplă semnificație ordinală, ci și una *cardinală*. O instanță a unei atribuiri de valoare cardinală este temperatura. Fie că măsurăm temperatura în grade Celsius, fie că o măsurăm în grade Fahrenheit, valoarea de adevăr a afirmației „temperatura medie în Paris este mai mare decât temperatura medie în New York" nu se modifică. (Dacă temperatura este măsurată ordinal, acest enunț nu ar avea sens.) Dimpotrivă, valoarea enunțului „Este de două ori mai cald în Paris decât în New York" *depinde* de alegerea scării de măsurare. Însă, cu toate că valoarea de adevăr a acestui enunț particular cu privire la intensitate este sensibilă la scara de măsurare aleasă, alte enunțuri nu sunt. Valoarea de adevăr a enunțului „Diferența de temperatură dintre New York și Paris este mai mare decât cea dintre Paris și Oslo", de pildă, nu depinde de alegerea scării de măsurare. Asemănător, putem construi măsuri cardinale de utilitate care reflectă – printre altele, după cum vom vedea – intensitatea preferințelor, și nu doar clasificarea ordinală a opțiunilor. Acestea ne dau posibilitatea să comparăm câștigurile (sau pierderile) de utilitate implicate de trecerea de la x la $(x + 1)$ cu trecerea de la $(x + 1)$ la $(x + 2)$, adică ne permite să vorbim despre utilitatea marginală în creștere sau în scădere – concepte care nu au sens pentru măsuri ordinale ale utilității.

Detaliile tehnice ale construcției nu trebuie să ne preocupe întrucât ideea de bază este simplă și suficientă pentru scopurile urmărite aici. Vom începe prin a presupune că agenții au preferințe nu doar pentru opțiuni, ci și pentru *loterii* de opțiuni (includzând aici „loteriile degenerate", care constau în obținerea unei opțiuni fundamentale în mod sigur). Pentru orice set dat de opțiuni fundamentale sau „prețuri", o loterie precizează, pentru fiecare preț, probabilitatea de a-l obține, suma tuturor probabilităților fiind egală cu 1. Se presupune că agenții au preferințe complete și tranzitive pentru astfel de loterii. De asemenea, se presupune că preferințele respectă „axioma independenței": preferința între două loterii p și q nu este afectată dacă sunt ambele combinate în același mod cu o a treia loterie r. „Efectul certitudinii", citat în capitolul 7 și discutat apoi în capitolul 12, nu respectă această axiomă.

În cele din urmă, se presupune că preferințele manifestă o formă de continuitate, definită după cum urmează. Să presupunem că opțiunile fundamentale includ un element care este cel mai bun, A, precum și un element care este cel mai rău, B. Acestora le atribuim în mod arbitrar valorile de utilitate 1 și 0. Continuitatea presupune că,

pentru orice opțiune intermediară C, există o probabilitate p(C) care l-ar face pe agent indiferent între a obține C în mod sigur și a intra într-o loterie care i-ar da A cu probabilitatea p(C) și B cu probabilitatea 1 – p(C).[1] Definim apoi *utilitatea cardinală* u(C) ca fiind egală cu p(C). Firește, acest număr este arbitrar, ca și utilitățile finale. Să presupunem că atribuim valorile de utilitate M și N lui A și, respectiv, lui B (M > N). Putem defini apoi utilitatea lui C drept utilitatea așteptată a loteriei:

$$pM + (1 - p)N = Mp + N - Np = (M - N)p + N.$$

Clasa funcțiilor de utilitate care apar astfel este mult mai mică decât clasa funcțiilor de utilitate ordinală.[2] Este lesne de observat că, dacă opțiunea X are o utilitate așteptată mai mare decât Y conform unei funcții, atunci va avea o utilitate așteptată mai mare conform altei funcții. Putem astfel susține fără ambiguitate că un agent rațional își maximizează utilitatea așteptată.

Funcțiile de utilitate cardinală au proprietatea foarte importantă de a fi *funcții de probabilitate liniare*. Să introducem notația XpY, care reprezintă o loterie ce oferă probabilitatea p pentru a obține X și 1 – p pentru a obține Y. Utilizând o scală care merge de la 0 la 1, utilitatea u(X) este egală cu probabilitatea q la care agentul este indiferent între X și loteria AqB. Asemănător, utilitatea u(Y) este egală cu probabilitatea r la care este indiferent între Y și loteria ArB. XpY, prin urmare, oferă echivalentul de utilitate al unei șanse p de a obține A cu probabilitatea q și o șansă 1 – p de a obține A cu probabilitatea r. Utilitatea lui XpY, prin urmare, este $pq + r(1 - p)$, care este p înmulțită cu utilitatea lui X plus (1 – p) înmulțită cu utilitatea lui Y. De pildă, utilitatea combinației probabilistice a 3/5 șanse de a obține X și a 2/5 șanse de a obține Y este $3/5q + 2/5r$.

S-ar putea ridica următoarea obiecție. Să presupunem că un fermier are de ales între două tipuri de recolte: soiul tradițional, care are șanse egale de a produce fie o recoltă bună, fie una mediocră, în funcție de vreme, și soiul modern, care are șanse egale de a produce o recoltă excelentă sau una foarte săracă. Să presupunem că utilitățile cardinale sunt 3 și, respectiv, 2 pentru recolta în stil vechi și 5 sau 1 pentru cea în stil modern. Din moment ce utilitatea așteptată pentru noua recoltă este mai mare, fermierul ar trebui să o aleagă pe aceasta. Dar, s-ar putea obiecta mai departe, oare nu se trece astfel cu vederea faptul că fermierul poate are o aversiune mare față de risc și nu este dispus să accepte orice opțiune care ar putea duce la un nivel de utilitate atât

---

[1] Identificarea probabilității ridică problema ancorării, pe care am menționat-o în Introducerea la Partea a II-a.

[2] Oricare două astfel de funcții sunt, de fapt, în legătură una cu alta, așa cum sunt scalele de temperatură Celsius și Fahrenheit, care atribuie valori diferite (corespunzând lui M și lui N din text) temperaturilor la care apa fierbe și, respectiv, îngheață.

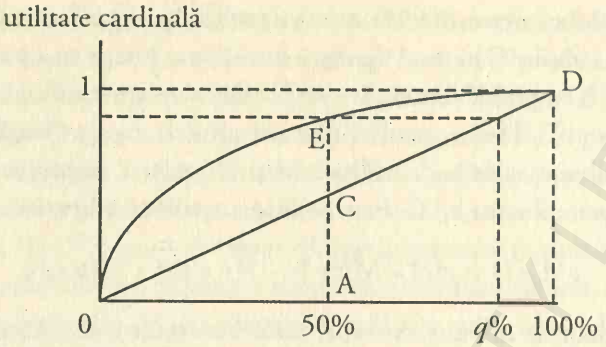

FIGURA 11.2

de mic precum 1? Obiecția presupune o numărătoare dublă, totuși, întrucât aversiunea față de risc este deja inclusă în construcția utilităților cardinale. Presupunând că A, B și C iau valorile 100, 0 și 60, u(C) ar putea foarte bine să fie 0,75 pentru o persoană cu aversiune față de risc, ceea ce implică indiferența între a obține 60 la sigur și o loterie care îl lasă cu 25% șanse de a nu obține nimic și 75% șanse de a obține 100. Un argument asemănător se aplică atribuirii de valori de utilitate cardinală cantităților fizice ale recoltei.

Pentru alt exemplu, putem lua în considerare atribuirea custodiei copilului (vezi Figura 11.2). Putem înțelege axa orizontală în două moduri: ca implicând fie o diviziune fizică a custodiei (procent din timpul petrecut cu copilul), fie o diviziune probabilistică (șansa de a obține custodia în întregime în urma unui proces). Utilitatea cardinală a împărțirii egale a timpului este AE, care este mai mare decât utilitatea AC a unei șanse de 50% a custodiei totale. (Facem aici apel la faptul că utilitatea cardinală este o funcție de probabilitate liniară.) Motivul este acela că, în astfel de cazuri, cei mai mulți oameni manifestă aversiune față de risc. Aceștia sunt dispuși să accepte custodia comună pentru că un risc de 50% de a nu-și vedea copilul deloc este intolerabilă. Doar în cazul în care unul dintre părinți crede că are o șansă mai mare de q% de a obține custodia în întregime este preferabil să recurgă la un proces decât să accepte custodia comună. Avem de a face cu o pondere considerabilă de procese de custodie nu pentru că părinții sunt iubitori ai riscului, ci pentru că gândirea deziderativă îi face să-și exagereze șansele de a obține custodia în totalitate.

Aversiunea față de risc și diminuarea utilității marginale

Deși corecte, lucrurile spuse până acum ar putea induce în eroare. Se manifestă o tendință într-o anumită zonă a literaturii de specialitate de a estompa distincția dintre aversiunea față de risc și diminuarea utilității marginale. Pentru a dezvolta acest argument, trebuie să introduc în discuție un concept care are sens din punct de vedere

intuitiv, cu toate că nu are (până acum) și o expresie măsurabilă. Este vorba despre ideea *utilității intrinseci* a bunului, care reflectă intensitatea preferințelor unui agent. Introspecția ne spune într-un mod convingător că anumite bunuri sau experiențe sunt extrem de plăcute, altele doar satisfăcătoare, altele ușor enervante, iar altele de-a dreptul oribile. Pentru a reprezenta diferența dintre acestea doar în termeni de preferințe ordinale – „Prefer Raiul Iadului, la fel cum prefer patru mere în loc de trei mere" – înseamnă să folosim o noțiune inadecvată a bunăstării sau utilității. Faptul că nu există nicio metodă sigură de a atribui numere nivelurilor intrinseci de satisfacție și insatisfacție nu dovedește că ideea în sine este lipsită de sens, la fel cum inabilitatea noastră de a cuantifica și de a compara nivelurile de satisfacție ale unor indivizi diferiți ne arată că ideea comparațiilor interpersonale ale bunăstării este lipsită de sens.

Ideea că multe bunuri au o utilitate marginală descrescătoare poate fi înțeleasă din această perspectivă. Pentru un sărac, primii dolari au o utilitate mare, dar apoi fiecare dolar își pierde din valoare la nivel subiectiv. Fiecare fumător știe că prima țigară de dimineață este cea mai bună și că te bucuri mult mai mult de țigări dacă fumezi mai rar. Fumatul unei țigări are, de fapt, două efecte: produce satisfacție în prezent și reduce satisfacția țigărilor fumate în viitor.

Însă cel de-al doilea efect nu trebuie să fie negativ. Să luăm din nou în considerare cazul custodiei copilului. Pentru un părinte, o după-amiază petrecută cu copilul la sfârșit de săptămână poate produce mai mult frustrare decât satisfacție. O după-amiază în fiecare sfârșit de săptămână este de două ori mai satisfăcătoare pentru că legăturile emoționale puternice create de întâlnirile mai dese le face pe fiecare mai satisfăcătoare. La celălalt capăt al spectrului temporal, satisfacția suplimentară obținută din faptul de a fi în prezența copilului șapte zile din săptămână mai degrabă decât șase depășește satisfacția suplimentară dată de șase zile, și nu de cinci, pentru că o custodie deplină asigură beneficiul lipsei de constrângere în planificare. Faptul de fi în prezența copilului are, de fapt, o utilitate marginală (intrinsecă) crescătoare, după cum reiese și din Figura 11.3.

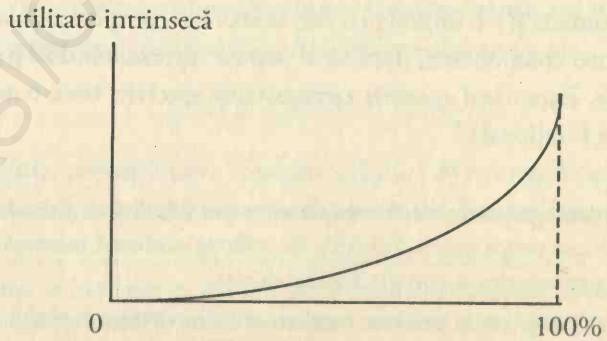

FIGURA 11.3

BCU IASI CLUJ LIBRARY

În acest caz, interpretarea axei orizontale este procentul de timp petrecut cu copilul. Pentru motivele date mai sus, fiecare oră în plus este mult mai valoroasă decât ora precedentă. *Acest enunț este perfect compatibil cu analiza subiacentă Figurii 11.2.* Utilitatea marginală a timpului petrecut cu copilul poate fi descrescătoare, dacă înțelegem utilitatea ca utilitate cardinală, și crescătoare, dacă o concepem în termeni de utilitate intrinsecă. Faptul că doar prima parte a acestui enunț are o interpretare măsurabilă nu înseamnă că a doua parte a enunțului este lipsită de sens.

În timp ce funcțiile de utilitate cardinală sunt întotdeauna generate de doi factori psihologici subiacenți, anume atitudinea față de risc și utilitatea intrinsecă, acestea nu pot fi măsurate separat. Nu putem spune într-un mod riguros dacă, de pildă, curba OED din Figura 11.2 este derivată din neutralitatea față de risc combinată cu o diminuare a utilității marginale intrinseci a timpului petrecut cu copilul sau din aversiunea față de risc combinată cu o creștere a utilității marginale intrinseci a timpului petrecut alături de copil. Într-un caz anume, intuiția ne poate spune care dintre cele două interpretări este mai plauzibilă. Pentru unii părinți, timpul petrecut alături de copil poate fi experimentat într-un mod asemănător cu acela al multor bunici: este bun în doze mici, dar devine epuizant în scurt timp. În același timp, este posibil ca acești părinți să nu se teamă deloc de riscul de a nu petrece timp alături de copil (neutralitate față de risc). Alți părinți pot fi diferiți din ambele puncte de vedere, generând aceeași funcție de utilitate cardinală. Insist asupra acestei chestiuni: aceste enunțuri nu pot fi prezentate (până la momentul de față) într-un mod riguros, dar, în mod evident, au sens.[1]

Credințe raționale

Cu aceasta încheiem discuția cu privire la prima componentă a alegerii raționale: a alege cele mai bune mijloace pentru a realiza dorințele, date fiind credințele. Firește, aceasta este doar o condiție necesară pentru raționalitate, nu una suficientă. Dacă vreau să-mi omor vecinul și cred că modul cel mai bun de a ucide un om este să fac o păpușă care îi seamănă și s-o împung cu ace, atunci mă comport rațional (din punctul de vedere al primei componente) dacă fac o păpușă reprezentându-l pe vecinul meu și o împung cu ace. Exceptând anumite circumstanțe speciale, însă, o astfel de credință este departe de a fi rațională.[2]

---

[1] Prin urmare, analogia cu scalele de temperatură este validă doar parțial. Aceste scale măsoară *doar* intensitatea temperaturii. Funcțiile de utilitate cardinală măsoară rezultatul conjugat al intensității preferințelor și atitudinilor față de risc.

[2] Credința în vrăjitorie poate produce rezultate reale în virtutea faptului că acestea au fost prezise, dacă persoana blestemată crede în eficiența blestemului și își pierde pofta de viață. Într-un astfel de caz, eficiența observată a blestemului poate face credința în vrăjitorie rațională,

Credințele raționale sunt acelea modelate de procesarea probelor disponibile prin utilizarea unor proceduri care, pe termen lung și în medie, au cea mai mare șansă de a produce credințe adevărate. Să presupunem că vrem să ne formăm o credință despre șansele să plouă pe 29 noiembrie, de acum într-o săptămână. Probabil că opțiunea cea mai bună este aceea de a căuta statistici despre precipitații din anii anteriori și de a presupune că viitorul (așteptat) va fi ca trecutul. Însă, pe măsură ce data de 29 noiembrie se apropie, precipitațiile curente ne pot face să ne schimbăm așteptările. Dacă plouă des în noiembrie și noi ne bucurăm în fiecare zi de cer senin, putem infera existența unui sistem de presiune atmosferică înaltă care diminuează semnificativ șansele de ploaie pe 29 noiembrie.

Acest proces de revizuire a credințelor este numit adesea *învățare bayesiană* (după pastorul Bayes, din secolul al XVIII-lea). Să presupunem că avem o distribuție de probabilitate subiectivă („inițială") peste diferite stări ale lumii. În exemplul dat mai sus, distribuția inițială a fost dedusă din frecvențele observate în trecut. În anumite cazuri, aceasta poate fi o simplă bănuială. Pe baza intuiției mele, pot atribui, de pildă, o probabilitate de 60% că prim-ministrul (PM) este competent și o probabilitate de 40% că acesta este incompetent. Putem observa măsurile pe care acesta le ia în timpul mandatului și consecințele lor, cum ar fi rata de creștere a economiei. Să presupunem că ne formăm o opinie despre șansele acestor observații *dată fiind competența* PM. Cu un PM competent avem o așteptare de 80% a unui rezultat bun, pe când cu unul incompetent avem doar 30%. Bayes a arătat modul în care putem actualiza probabilitățile inițiale privitoare la competența PM, *date fiind observațiile pe care le facem.*

Să presupunem că există doar două rezultate posibile, bun sau rău, și că observăm un rezultat bun. Dacă notăm cu $p(a)$ probabilitatea să rezulte a și cu $p(a \mid b)$ *probabilitatea condiționată* să rezulte a, dat fiind că rezultă b, am asumat că p(PM este competent) = 60%, p(PM este competent) = 40%, p(rezultat bun | PM este competent) = 80% și p(rezultat bun | PM este incompetent) = 30%. Căutăm să determinăm valoarea p(PM este competent | rezultat bun). Folosim literele a și b pentru a denota competența și, respectiv, rezultatul bun. Apoi facem observația că

$$p(a \mid b) = p(a \,\&\, b) / p(b) \qquad (*)$$

Cu alte cuvinte, probabilitatea condiționată $p(a \mid b)$ este egală cu probabilitatea să rezulte atât a, cât și b, împărțită la probabilitatea lui b. Aceasta decurge din ideea mai

---

chiar dacă (așa cum se întâmplă cu teoria acțiunii la distanță) agentul nu poate preciza mecanismul prin care funcționează. O poate face chiar și pasibilă de pedepse pe baza consecințelor sale reale și nu, cum au sugerat Donne și Hobbes (vezi Introducerea la Partea a II-a), doar în baza existenței unei *mens rea*.

intuitivă că $p(a \,\&\, b)$ este egal cu $p(b)$ înmulțit cu $p(a \mid b)$. Împărțind ambele părți ale ecuației la $p(b)$, obținem ecuația (*).

Folosind ecuația (*) din nou, dar cu a și b inversate, avem

$$p(b \mid a) = p(a \,\&\, b) \,/\, p(a)$$

sau, echivalent

$$p(a \,\&\, b) = p(b \mid a) \cdot p(a)$$

Substituind ultima expresie în (*), obținem

$$p(a \mid b) = p(b \mid a) \cdot p(a) \,/\, p(b) \qquad (**)$$

Acum, există două moduri pentru ca b (rezultatul bun) să apară, cu un PM competent sau cu un PM incompetent. Făcând apel la faptul că probabilitatea ca unul din două evenimente mutual exclusive să se producă reprezintă suma probabilităților fiecărui eveniment, putem astfel scrie

$$p(b) = p(b \,\&\, a) + p(b \,\&\, \text{non-}a)$$

care, în urma raționamentului din paragraful care urmează (*), este echivalentă cu

$$= p(b \mid a) \cdot p(a) + p(b \mid \text{non-}a) \cdot p(\text{non-}a)$$

Dacă substituim expresia pentru $p(b)$ în (**), obținem *teorema lui Bayes*:

$$p(a \mid b) = p(b \mid a) \cdot p(a) \,/\, [p(b \mid a) \cdot p(a) + p(b \mid \text{non-}a) \cdot p(\text{non-}a)]$$

Înlocuind valorile numerice de probabilitate în partea dreaptă a ecuației, rezultă că $p(a \mid b) = 80\%$, adică observarea unui rezultat bun crește șansele ca PM să fie competent de la 60% la 80%. O a doua și o a treia observație pozitivă ar ridica șansele la 91%, iar apoi la 97%.[1] Dacă altă persoană a estimat inițial $p(a) = 0{,}3$, și nu 0,6, trei

[1] În acest raționament ne bazăm pe credințe inițiale de două tipuri, o „credință de prim-plan", care vizează tipul de PM, și o „credință de fundal", cu privire la șansele ca un PM de un anumit tip să producă un rezultat bun. În timp ce credința din fundal rămâne constantă tot timpul, credința din prim-plan este revizuită pe măsură ce primim noi informații. Putem chiar să mutăm prim-planul în fundal și invers, în cazul în care ne interesează dacă politicienii competenți sunt, de fapt, capabili să modeleze rezultate economice sau dacă acestea sunt afectate în mare parte de întâmplare și de factori exogeni. Putem apoi începe cu o credință de prim-plan despre șansele ca politicienii buni să producă rezultate bune și cu o credință de fundal despre competența unui anumit politician. Dacă observațiile arată că economia o duce rău sub cârmuirea unui politician despre care am presupus că va avea o șansă mare de a fi un tip competent, atunci putem atribui șanse mai mari ipotezei referitoare la influențele întâmplătoare sau exogene.

observații pozitive succesive ar ridica estimarea mai întâi la 0,53, apoi la 0,75 și, în cele din urmă, la 0,89. Prin urmare, se poate să nu conteze prea mult faptul că bănuielile inițiale nu sunt sigure din moment ce, pe măsură ce crește nivelul de informație, credințele actualizate devin din ce în ce mai sigure. În timp, diferențele inițiale de opinie pot fi înlocuite de noi probe.[1] Ceva mai târziu (capitolul 23), vom observa că fiecare informație nouă are un impact mai mic decât cea anterioară.

Investiția optimă în culegerea de informații

A treia componentă a acțiunii raționale este investiția optimă în resurse – precum timpul și banii – pentru dobândirea de informații. După cum reiese din Figura 11.1, există mai multe elemente determinante ale acestui optim. În primul rând, cantitatea de informație necesară depinde de dorințele agentului.[2] De pildă, un agent căruia nu îi pasă prea mult de recompense în viitorul îndepărtat nu va investi prea mult în determinarea duratei de viață așteptate a unui bun de consum durabil. Și, mai evident, are sens să culegi mai multe informații înainte de a lua o decizie importantă, cum ar fi cumpărarea unei case, ceea ce nu este același lucru cu alegerea între două sticle de vin la fel de scumpe. În cel din urmă caz, cel mai bun lucru ar fi, probabil, luarea unei decizii prin aruncarea unei monede, dacă costul așteptat al hotărârii cu privire la care sticlă este mai bună depășește beneficiile așteptate (pe baza unor cunoștințe prealabile despre nivelul calității vinurilor care au același preț) ale consumării sticlei mai bune de vin, mai degrabă decât a aceleia inferioare calitativ.

Dorințele și credințele anterioare determină împreună beneficiile așteptate din obținerea de informații noi. Uneori este posibil să aflăm cu precizie foarte mare numărul suplimentar de vieți care vor fi salvate în urma unui test specific pentru cancer sau, dacă ducem lucrurile la nivelul agentului, cât este de probabil ca viața să îi fie

B.C.U. „M. EMINESCU" IAȘI

---

[1] Pentru a apărea convergența, informațiile noi succesive trebuie să fie independente unele de altele din punct de vedere statistic. În exemplul de manual al modului în care se formează credințe pe model bayesian, o persoană extrage bile dintr-o urnă despre care se știe că are, în egală măsură, probabilitatea să conțină fie 80% bile negre și 20% albe, fie 20% negre și 80% albe pentru a determina care sunt șansele obținerii unuia dintre rezultate. Din moment ce extragerile sunt întâmplătoare și bilele sunt puse din nou în urnă după fiecare extragere, rezultatul fiecărei extrageri este independent de cel precedent. În situații politice precum cea descrisă în text, independența poate fi mult mai greu de verificat. De asemenea, convergența presupune că situația subiacentă rămâne aceeași sau, cel puțin, că nu se modifică prea repede. Dacă mediul se schimă foarte rapid, procesul de actualizare a opiniilor seamănă cu acela de țintire a unui obiect în mișcare (capitolul 6).

[2] Gândirea dezideративă, în cazul căreia „dorința este mama gândului", este clar irațională. Nu este, însă, nimic irațional în legătură cu procesul ilustrat prin Figura 11.1, în care dorințele sunt, cum ar veni, bunicile credințelor.

salvată. Valoarea vieții depinde de modul în care agentul înțelege să compenseze viața în raport cu alte scopuri dorite. În urma unor anumite calcule, este nevoie de un bonus de vreo 200 de dolari pe an pentru a-i face pe cei care au ocupații riscante, precum munca în mine de cărbune, să accepte un risc de o mie pe an de moarte accidentală. Prin urmare, în momentul în care s-a făcut acest calcul, valoarea vieții era undeva la 200 000 de dolari.[1] Costurile așteptate ale noilor informații, care sunt determinate de credințe anterioare, pot fi uneori evaluate cu precizie. Pentru detectarea cancerului intestinal se obișnuiește efectuarea a șase teste de scaun ieftine. Beneficiile primelor două teste sunt semnificative. Însă, pentru fiecare dintre celelalte patru teste, costurile detectării unui caz de cancer în plus (pe care nici măcar nu îl vindecă) sunt 49 150 de dolari, 469 534 de dolari, 4 724 695 de dolari și 47 107 214 de dolari.

Căutarea optimă de informații poate să depindă și de rezultatele căutării înseși (bucla din Figura 11.1 reflectă acest lucru). Când este testat un nou produs medical, există o decizie anterioară de a administra medicamentul unui grup și de a-l reține pentru alt grup pentru o anumită perioadă de timp. Însă, dacă se demonstrează de la bun început că produsul este spectaculos de bun, nu ar fi etic să nu îl administrezi și grupului de control. Același argument se poate aplica și unui singur agent rațional. Să presupunem că merg la cules de fructe de pădure. Știu că acestea cresc de obicei în grupuri, prin urmare sunt pregătit să caut ceva timp înainte să mă apuc de cules. Dacă am noroc și găsesc o zonă bună de la început, ar fi absurd să caut în continuare.

Putem socoti culegerea de informații drept o *acțiune din umbră* care însoțește acțiunea primară. Înainte să decidem ce e de făcut trebuie să decidem cât de multe informații să culegem. Uneori, *acțiunea din umbră și acțiunea primară* pot coincide, cel puțin parțial. Să presupunem că liderii unei țări iau în calcul posibilitatea de a merge la război împotriva altei țări. Invadarea Franței de către Germania în 1940 poate servi ca exemplu. Pentru a lua o decizie cu privire la atac, informația era crucială. Trebuia ca liderii să știe care erau capacitățile obiective și potențiale ale inamicului, cum ar fi „organizarea, cutumele și obiceiurile armatei inamice" (din manualul german *Datoriile Înaltului Comandament*). Mare parte din informație putea fi culeasă prin mijloace convenționale, inclusiv spionajul. Totuși, pentru a determina moralul inamicului – spiritul combatant – nu exista alternativă în afara luptei.

Indeterminarea

Aceste ultime exemple – culesul de fructe de pădure și planificarea unui război – ne vor ajuta să vedem *limitările* teoriei alegerii raționale sau, mai degrabă, una dintre

[1] Există multe capcane implicate de astfel de calcule, dar argumentul central este imposibil de respins: cu toții acordăm o valoare finită vieților noastre. Dacă nu ar fi așa, nu ne-am angaja deloc în toate activitățile plăcute sau profitabile care presupun risc.

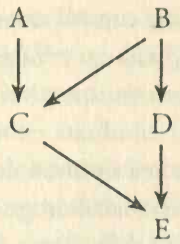

FIGURA 11.4

cele două limitări. Ca instrument explicativ, teoria poate eșua într-una din două privințe. Pe de o parte, poate eșua în a produce predicții unice cu privire la ceea ce vor face oamenii într-o anumită situație. Pe de altă parte, oamenii pot eșua în a fi pe măsura predicțiilor, fie acestea unice sau nu. Al doilea eșec, iraționalitatea, este subiectul capitolului următor. Primul eșec, indeterminarea, este subiectul următoarelor observații.

Un agent poate să nu fie în măsură să identifice elementul cel mai bun dintr-un set fezabil, pentru unul din două motive. Un consumator poate fi *indiferent* între două opțiuni care sunt egal și maximal de bune. În cazurile triviale, acest lucru se întâmplă când opțiunile nu pot fi distinse, de exemplu, în cazul în care un consumator are de ales între două conserve de supă la supermarket. În cazurile netriviale, două opțiuni pot să difere în funcție de mai mulți parametri, astfel încât diferențele se află într-un echilibru perfect. Cazurile netriviale sunt rare, poate inexistente. Dacă am de ales între două mașini diferite din punctul de vedere al prețului, confortului, aspectului, vitezei și așa mai departe, nu pot să o prefer pe vreuna în defavoarea celeilalte fără să fiu, de asemenea, indiferent între opțiuni. În caz contrar, o reducere de cinci cenți la o mașină ar trebui să determine o preferință pentru acea opțiune. Intuiția sugerează că este puțin probabil ca așa ceva să se întâmple.

De fapt, preferințele consumatorilor pot fi *incomplete*. Să presupunem că am inspectat cinci modele de mașini, A, B, C, D și E, și că le ordonez ca în Figura 11.4 (săgețile reprezintă relația de preferință). Incapacitatea mea de a compara C cu D nu contează din moment ce oricum nu am de gând să o cumpăr pe vreuna dintre cele două. Dimpotrivă, inabilitatea mea de a compara A și B îmi creează o problemă. E adevărat, pot încerca să culeg mai multe informații, dar de unde știu că merită efortul? Mă voi întoarce în curând la acest aspect. În primul rând, însă, voi sublinia o altă sursă, probabil mai importantă, de preferințe incomplete. În mod obișnuit, preferințele pentru anumite opțiuni sunt induse de preferințele rezultatelor. Prefer o anumită opțiune pentru că prefer rezultatul acesteia, adică utilitatea sa așteptată, în comparație cu alte opțiuni. Însă, dacă situația este una de incertitudine sau ignoranță, și nu una

de risc (capitolul 7), se poate să nu fiu capabil să compar rezultatele.[1] În cuvintele nemuritoare ale lui Samuel Johnson, „viața nu e lungă și nu trebuie să petreci prea mult timp în deliberări inutile despre cum anume o vei trăi: aceia care încep această deliberare cu prudență și o continuă cu subtilitate trebuie, după risipirea multor gânduri, să o încheie prin întâmplare. A prefera un mod de viață viitor bazându-ne pe temeiuri motivate în locul altuia presupune facultăți pe care Creatorul nu ni le-a hărăzit."[2]

O altă indeterminare provine din dificultatea de a determina investiția optimă în culegerea de informații. Când culeg fructe de pădure, cât timp ar trebui să caut un grup compact de plante și când ar trebui să mă apuc, efectiv, de cules? Cu excepția cazului în care găsesc de la bun început o zonă bogată, are sens să petrec o vreme uitându-mă în jur. În același timp, nu vreau să pierd vremea până seara pentru că atunci sigur mă voi duce acasă cu coșul gol. Între limita superioară și inferioară de timp poate exista un interval destul de mare de indeterminare. Ne confruntăm cu o problemă diferită în cazul planificării unui război. Dacă decizia primară și decizia din umbră coincid, planificatorul este sortit să rămână într-o stare de incertitudine (cel puțin parțială). Teoria alegerii raționale nu poate să ne ghideze în astfel de situații. Ea este de ajutor în situații foarte bine structurate, despre care se știe destul de mult, cum ar fi testele pentru cancer, însă nu este foarte utilă în medii necunoscute.

Deși faptul de a petrece mai puțin timp decât limita inferioară sau mai mult timp decât limita superioară ar putea fi comportamente iraționale, nicio alegere pe care o face agentul în cadrul acestui interval nu poate fi caracterizată ca irațională. Prin urmare, ne-am putea gândi să renunțăm la ideea de raționalitate în favoarea aceleia de *neraționalitate*. Această versiune revizuită a teoriei ne-ar permite să înțelegem o gamă mai largă de comportamente, dar are puteri predictive mai mici. Cei mai mulți avocați ai teoriei, cred, ar fi reticenți să o revizuiască în acest mod. Ce îi atrage la teorie este faptul că promite să genereze predicții *unice*. O face prin faptul matematic elementar că orice funcție de utilitate normală definită pe un set de oportunitate normal

---

[1] În luarea de decizii referitoare la situații de incertitudine, pot fi în măsură să compar opțiunile, în cazul în care cel mai rău rezultat al unei opțiuni este mai bun decât cel mai bun rezultat al altei opțiuni. În luarea de decizii în situații de ignoranță, chiar și această comparație modestă nu este disponibilă.

[2] Pot, însă, „conchide prin șansă", iar apoi pot inventa „motivele temeinice", de exemplu, acordând o pondere mai mare acelor atribute pentru care opțiunea aleasă este superioară. Acest lucru poate avea consecințe nedorite. Să presupunem că am de ales între a studia dreptul și a studia silvicultura. Nefiind în stare să fac o alegere fundamentată pe rațiune, merg la drept, mai mult sau mai puțin din întâmplare, și justific decizia retrospectiv acordând o pondere mai mare aspectului financiar presupus de cele două cariere. Cu aceste noi preferințe induse, pot continua prin a lua alte decizii, care diferă de acelea pe care le-aș fi luat pe baza preferințelor mele de dinaintea alegerii inițiale.

atinge o valoare maximă pentru un membru unic al acelui set. Interacțiunea dintre setul de oportunitate și curbele de indiferență din Figura 9.1 ne oferă un bun exemplu al simplității foarte convingătoare al ideii de „a face cel mai bun lucru pe care ești în măsură să îl faci".

Cea mai importantă sursă a indeterminării credințelor se manifestă în interacțiunea strategică, atunci când fiecare agent trebuie să își formeze credințe despre ce este probabil că vor face ceilalți pe baza credințelor *lor*, știind că aceștia trec prin raționamente similare cu privire la credințele proprii ale agentului. În unele cazuri, pe care le vom aborda în capitolul 19, structura recompensei nu le permite agenților să conveargă spre un set comun de credințe.

Raționalitatea este în întregime subiectivă

Voi încheia discuția despre teoria deciziei raționale scoțându-i din nou în evidență *natura radical subiectivă*. Firește, am putea considera cuvântul „rațional" într-un sens obiectiv, înțelegând prin aceasta faptul că agentul ia decizii care *îi fac viața mai bună* judecând după criterii obiective, cum ar fi sănătatea, longevitatea sau venitul. Dacă am folosi-o astfel, însă, ideea nu ar avea niciun fel de putere explicativă. După cum am subliniat deja, *consecințele* unei decizii nu pot explica acea decizie. Doar stările mentale care o precedă ne dau posibilitatea să *explicăm* acțiunile ca fiind optime din punctul de vedere al agentului, nu să le *caracterizăm* drept utile sau benefice din punctul de vedere al unui observator extern (sau al agentului la un moment ulterior).

Să presupunem că sufăr de incapacitatea extremă de a amâna recompensele, adică nu sunt în stare să iau în considerare consecințele viitoare ale comportamentului meu actual. Și să presupunem că oamenii de știință ar produce o pastilă împotriva deprecierii timpului, care ar crește ponderea recompenselor viitoare ale unor decizii luate în prezent. Dacă voi lua pastila, viața mea va deveni mai bună. Părinții mei vor fi mai fericiți pentru că am luat pastila. Privind retrospectiv, voi fi bucuros pentru ce am făcut. Însă, în cazul în care am de ales să iau pastila sau nu, voi refuza dacă sunt rațional. Orice comportament pe care l-ar induce pastila este, deja, la îndemână. Aș putea să mă las de fumat, să mă apuc de exerciții fizice sau să pun bani deoparte, dar nu fac nimic din toate acestea. Din moment ce nu vreau să fac aceste lucruri, nu voi vrea să iau pastila care m-ar constrânge să le fac. Similar, o persoană egoistă ar refuza o „pastilă a altruismului" și, chiar mai convingător, un altruist ar refuza o „pastilă a egoismului". Dacă îmi iubesc familia și sunt dispus să sacrific o parte din bunăstarea hedonistă de dragul lor, voi refuza o pastilă care, în două etape, ar avea ca efect o diminuare a bunăstării lor, la fel cum aș refuza orice opțiune (de exemplu, să cumpăr o masă foarte scumpă pentru mine), care ar produce același efect într-o singură etapă.

Putem rafina argumentul presupunând că persoana consumă x astăzi și y mâine și că rata deprecierii per perioadă este 0,5 (persoana este indiferentă între o unitate de

utilitate mâine și o jumătate de unitate astăzi). Să presupunem, pentru simplitate, că $u(x) = x$ și $u(y) = y$. Valoarea depreciată prezentă a fluxului de consum este $x + 0{,}5y$. Să presupunem că persoana află că mâine urmează să sufere de o anumită durere care îi va reduce utilitatea consumului cu un factor 0,5. Valoarea depreciată prezentă a consumului este acum $x + 0{,}25y$. Dacă unui agent rațional i se oferă gratuit o pastilă de aspirină care va elimina durerea, persoana ar trebui să accepte oferta, restabilind astfel valoarea prezentă inițială. Dacă ar lua o pastilă care ar determina o rată a deprecierii de 1 (dar nu ar lua aspirina), rezultatul ar fi același în sensul că, în ambele cazuri, persoana ar fi indiferentă între fluxul constituit din două perioade și o utilitate pentru o singură perioadă egală cu $x + 0{,}5y$. Din moment ce agentul ar trebui să ia aspirina și din moment ce efectele acesteia sunt aceleași ca ale pastilei deprecierii, de ce nu ar lua-o și pe cea din urmă? Motivul este acela că alegerea pastilei deprecierii este constrânsă de necesitatea superiorității consumului indus de pastilă în raport cu consumul fără pastilă, așa cum sunt acestea *judecate prin preferințele anterioare administrării pastilei*. În cazul aspirinei, nu avem de-a face cu constrângeri asemănătoare pentru că nu există nicio diferență între preferințele anterioare și cele posterioare administrării pastilei. Chiar fără aspirină, prefer să fiu în măsură să îmi asigur evitarea durerii mâine. Când acea stare devine parte din natura mea, aleg să o fac să apară. Dimpotrivă, fluxul de utilitate indus de pastila deprecierii face deja parte din natura mea, dar aleg să nu îl fac să se manifeste.[1]

Cu alte cuvinte, alegerile nu trebuie să fie văzute prin ochii agentului. Este posibil ca o persoană cu miopie care își pierde ochelarii să nu și-i găsească tocmai pentru că o împiedică boala. Persoana este „în capcană".[2] Asemănător, un agent rațional se poate găsi într-o „capcană a credințelor", care îl blochează la nivelul unei credințe fal-

---

[1] În cazul deprecierii hiperbolice, un agent poate accepta o pastilă a deprecierii. Folosind exemplul numeric din capitolul 6, putem presupune că efectul pastilei este acela de a diminua valoarea lui k de la 1 la 0,3. În momentul în care devine disponibilă recompensa mai mică, valoarea sa prezentă este doar 10 (nu avem depreciere). Valoarea prezentă a recompensei mai mari la acel timp este $30/(1+0{,}3\times 5) = 12$. Prin urmare, un preangajament sub forma înghițirii unei pastile îi va permite agentului să acționeze în conformitate cu judecata sa calmă și bine cumpănită, prevenind astfel slăbiciunea voinței (în sens larg). Acest enunț își păstrează valoarea de adevăr chiar dacă agentul trebuie să cumpere pastila, cât timp costul său (în termeni de utilitate) este mai mic de 2. Enunțul rămâne, de asemenea, adevărat dacă preangajamentul are ca efect diminuarea valorii recompensei întârziate (poate că pastila deprecierii are ca efect secundar reducerea capacității de a simți satisfacție), cât timp costul este mai mic decât 5. Aceste fapte ar putea fi relevante dacă înlocuim pastila deprecierii prin psihoterapie.

[2] Dacă i s-ar oferi ochelarii, și i-ar pune la ochi. Am argumentat că, dacă agentului i s-ar oferi pastila deprecierii, acesta nu ar lua-o. Diferența este aceea că agentul poate deja să facă fără pastilă tot ce ar putea face cu ajutorul ei, însă sunt multe lucruri pe care nu le poate face decât dacă își pune ochelarii.

se, în cazul în care costurile pe care agentul crede că le presupune verificarea credinței sunt prea mari. Astfel, femeile care practică mutilarea organelor genitale se găsesc, probabil, într-o capcană a credințelor. Indigenii tribului Bambara din Mali cred că un bărbat va muri dacă, în timpul actului sexual, penisul lui atinge clitorisul. În Nigeria există triburi care cred că, dacă clitorisul atinge capul copilului în timpul nașterii, copilul va muri. În Polonia exista o credință larg răspândită că oricine consumă băuturi alcoolice în timp ce folosește disulfiram implantat subcutanat va muri. De fapt, disulfiramul implantat este inert din punct de vedere farmacologic. Însă falsa credință poate descuraja oamenii să o testeze.

Raționalitatea credințelor este o chestiune complet diferită de *adevărul* lor. În timp ce adevărul este o trăsătură a relației dintre credințe și lume, raționalitatea este o trăsătură a relației dintre credințe și probele deținute de agent. Cu toate că raționalitatea poate presupune ca un agent să investească în dobândirea de informații noi, investiția este constrânsă întotdeauna de costurile și beneficiile ei așteptate (adică acelea pe care agentul le *crede*). Dacă se crede că adunarea de informații suplimentare are costuri de *oportunitate* mari, cum ar fi, de pildă, în fața unui pericol posibil iminent, atunci poate fi rațional să refuzi investiția. Dacă se crede că acestea au costuri *directe* mari, precum în verificarea credințelor despre efectele consumului de alcool în timpul folosirii de disulfiram implantat, doar un agent irațional poate face o astfel de investiție. Într-un sens mai general, multe credințe trebuie luate în considerare așa cum par doar în mod secundar, din moment ce, dacă ar fi să le verificăm pe toate, nu ne-am mai putea trăi viețile.

Orice explicație a comportamentului bazată pe alegere este subiectivă. Însă nu toate explicațiile subiective pornesc de la presupoziția transparenței agenților față de ei înșiși și a căutării asidue a optimalității, care sunt caracteristici specifice ale explicațiilor prin alegerea rațională. În capitolul următor voi trece în revistă o serie de explicații ale căror caracteristici se disting parțial sau în totalitate de acelea ale teoriei alegerii raționale.

◈◈◈

Notă bibliografică

Am discutat relația dintre rațiune (în sensul din capitolul 4) și raționalitate în prelegerea mea inaugurală de la Collège de France, *Raison et raisons* (Paris, Fayard, 2006). Pentru detalii cu privire la concepția lui Weber despre raționalitate, vezi eseul meu „Rationality, economy, and society", în S. Turner (ed.), *The Cambridge Companion to Weber* (Cambridge University Press, 2000). O expunere clasică a teoriei utilității se regăsește în R.D. Luce și H. Raiffa, *Games and Decisions* (New York, Willey, 1957).

Lucrarea originală a lui J. von Neumann și O. Morgenstern, *The Theory of Games and Economic Behavior*; ediția a doua (Princeton, NJ, Princeton University Press, 1947) este încă de actualitate. O expunere excepțională a teoriei alegerii raționale (și problemelor sale) se află în R. Hastie și R. Dawes, *Rational Choice in an Uncertain World* (Thousand Oaks, CA, Sage, 2000). Am abordat pe larg exemplul legat de custodia copilului în capitolul 3 din *Solomonic Judgments* (Cambridge University Press, 1989). O prezentare elementară excelentă a teoriei bayesiene se regăsește în R. Winkler, *An Introduction to Bayesian Inference and Decisions* (Gainesville, FL, Probabilistic Publishing 2003). Argumentul meu că o persoană rațională nu ar lua pastila deprecierii a fost influențat și de unele schimburi de idei pe care le-am avut cu Gary Becker și Peter Diamond; vezi și O.J. Skog, „Theorizing about patience formation: The necessity of conceptual distinctions", *Economics and Philosophy*, 17 (2001), pp. 207-219. Am preluat ideea de capcană a credințelor de la G. Mackie, „Ending footbiding and infibulation: A convention account", *American Sociological Review*, 61 (1996), pp. 999-1017. Un studiu util privitor la importanța serviciilor de informații în pregătirea unui război este E.R. May, *Strange Victory: Hitler's Conquest of France* (New York, Hill & Wang, 2000). Datorez informația despre folosirea implanturilor de disulfiram în Polonia lui W. Osiatynski, *Alcoholism: Sin or Disease?* (Varșovia, Stefan Batory Foundation, 1997), iar datele despre ineficiența respectivului tratament lui J. Johnsen și J. Mørland, „Depot preparations of disulfiram: Experimental and clinical results", *Acta Psychiatrica Scandinavica*, 86 (1992), pp. 27-30.

BCU IAȘI/CENTRAL UNIVERSITY

Capitolul 12

RAȚIONALITATE ȘI COMPORTAMENT

Ignorarea costurilor deciziei

Ideea de raționalitate are o atracție normativă foarte puternică. *Vrem* să avem motive – dorințe și credințe în lumina cărora acțiunile noastre apar ca raționale – pentru ceea ce facem. De fapt, dorința noastră de a acționa dintr-un motiv – deferența noastră în fața raționalității – poate fi atât de puternică, încât determină un comportament irațional.[1] Putem defini hiperraționalitatea drept înclinația de a căuta decizia optimă din punct de vedere abstract, altfel spus, decizia care ar fi optimă dacă ar fi să ignorăm costurile pe care le presupune procesul luării de decizii în sine. Aceste costuri sunt de trei tipuri: (1) costul mijloacelor de a lua o decizie, (2) costurile efectelor secundare ale deciziei, (3) costurile de oportunitate ale deciziei, adică valoarea celorlalte lucruri pe care agentul le-ar fi putut face în loc să treacă prin întregul proces al luării unei decizii. Voi ilustra aceste puncte pe scurt.

Hiperraționalitatea prin neglijarea punctului (1) poate apărea în cazul acelora care fac *cumpărături prin comparație*, pentru care economisirea (așteptată) derivată din găsirea celui mai mic preț este mai mică decât banii cheltuiți pe transportul necesar deplasării dintr-un magazin în altul. Turiștii din sudul Franței trec granița cu Spania pentru a cumpăra țigări mai ieftine de parcă benzina ar fi gratis. Neglijarea punctului (2) poate determina hiperraționalitate în cazurile de custodie contestate. Tribunalul poate încerca să promoveze interesele copilului stabilind care dintre părinți este cel

[1] Nu ar trebui să confundăm acțiunea dintr-o *anumită* rațiune cu acțiunea conformă cu rațiunea, noțiune pe care am definit-o în capitolul 4. Persoana care alege în conformitate cu interesul său acționează dintr-o rațiune, dar nu în conformitate cu rațiunea. Invers, persoana care acționează conform imperativului categoric acționează în conformitate cu rațiunea, dar nu dintr-o rațiune (vezi nota 1, p. 87 a respectivului capitol). Cititorii care consideră că ultimul enunț este contraintuitiv au, probabil, o înțelegere diferită de a mea cu privire la semnificația ideii de a „acționa pentru o rațiune".

BCU LIBRARY

mai potrivit pentru încredințarea custodiei.[1] Odată stabilită chestiunea, tribunalul are un motiv temeinic pentru a încredința custodia părintelui respectiv. Însă procesul juridico-psihologic al evaluării adecvării părintelui poate avea consecințe dăunătoare incalculabile asupra copilului. O procedură mai rațională, dat fiind scopul care trebuie atins, ar fi să se recurgă la aruncarea unei monede sau să se rămână la presupoziția tradițională că mama este cea mai potrivită.

Neglijarea costurilor de oportunitate poate fi ilustrată printr-o observație a lui Samuel Johnson, emisă într-o conversație cu Boswell, despre materiile care ar trebui predate în primul rând copiilor: „Domnule, nu contează ce îi înveți prima dată, la fel cum nu contează care este piciorul pe care îți tragi mai întâi nădragii. Domnule, putem să ne certăm cu privire la piciorul care ar trebui să intre primul în nădragi, dar între timp vei sta dezbrăcat. În timp ce stai să te gândești cu care dintre două materii ar trebui să începi să-ți educi copilul, un altul deja le-a învățat pe amândouă."[2] Din nou, recursul la aruncarea unei monede poate fi cel mai rațional mod de a decide. Putem să ne gândim, de asemenea, la un doctor care ajunge la locul unui accident și trebuie să ia o decizie cu privire la pașii pe care îi are de urmat. Cu toate că trebuie, în mod evident, să examineze mai întâi pacientul, comportamentul său poate să fie contrar propriilor sale interese dacă stă prea mult timp să se gândească la ce are de făcut întrucât pacientul poate să moară între timp („Operația a fost un succes, dar pacientul a murit."). Alții, când au pornit la cules fructe de pădure, au trecut prin experiența de a căuta atât de mult locul potrivit, încât atunci când l-au găsit, se făcuse deja noapte. Chiar și atunci când economiile realizate în urma cumpărăturilor comparație sunt comparativ mai mari decât costurile de transport, comportamentul tot poate fi irațional dacă luăm în considerare valoarea timpului pierdut.

Câteva principii canonice ale raționalității

Prin „problemă paradoxală" voi înțelege acel comportament observat care pare impermeabil la explicațiile prin alegerea rațională. Cu toate că unele astfel de probleme, la o cercetare mai atentă, își pierd caracterul problematic, cele mai multe nu și-l pierd. Experimentele, precum și situațiile din viața reală, prezintă numeroase instanțe de modele comportamentale care violează canoanele raționalității. În cele ce urmează, din lista selectivă de astfel de principii canonice voi începe cu cele mai fundamentale și îmi voi croi drum către cele mai specifice. Mă voi limita doar la alegerile individuale:

---

[1] „Cel mai bun interes al copilului" este, de fapt, criteriul folosit în cele mai multe cazuri de custodie.

[2] Johnson, fiind un expert în Shakespeare, se poate să fi avut în minte replica regelui din *Hamlet* (III, 3): „Și ca un om aflat între ispite potrivnice nu-ncep nimic și preget, din două n-alegând niciuna."

anomaliile din alegerile interactive vor fi discutate în Partea a V-a. Lista poate fi completată cu unele dintre anomaliile cognitive pe care le-am discutat în capitolul 7.

1. În cazul unei alegeri între a acționa și a nu face nimic, un agent rațional nu va acționa dacă costurile de utilitate așteptate ale acțiunii depășesc beneficiile așteptate de utilitate.
2. În cazul alegerii dintre două rele, un agent rațional va alege răul mai mic.
3. Un agent rațional atribuie aceleași ponderi costurilor de oportunitate și costurilor directe.
4. Un agent rațional nu va prefera niciodată să aibă la dispoziție un subset dintr-un set de opțiuni în locul întregului set.
5. Dacă un agent rațional preferă X unui pahar descris ca fiind pe jumătate plin, agentul ar trebui, de asemenea, să prefere X unui pahar descris ca fiind pe jumătate gol.
6. Într-un joc pur al întâmplării, un parior rațional, atunci când face un pariu, nu va ține cont de rezultatele precedente ale jocului.
7. Când ia decizia dacă să rămână sau nu într-un proiect, un investitor rațional va ține cont doar de valoarea prezentă a fluxurilor viitoare de utilitate pe care le au cele două opțiuni.
8. Dacă la momentul 1 un agent rațional își face un plan de a duce la îndeplinire acțiunea X la momentul 2, atunci acesta va face X la momentul 2, chiar dacă dorințele și credințele sale s-au schimbat între timp.
9. În cazul unei alegeri riscante, un agent rațional va alege mijloacele în conformitate cu rezultatele așteptate, nu doar în conformitate cu scenariul cel mai bun (sau cel mai rău).
10. Într-o piață a agenților rationali, rata randamentelor tuturor activelor curente ar trebui să fie (aproximativ) aceeași.[1]
11. Dacă un agent rațional alege A dintr-un set (A, B, C), atunci acesta va trebui să aleagă A dintr-un set (A, B).
12. Un agent rațional nu va acționa asupra unui efect pentru a elimina cauza (va lua antibiotice, nu aspirină, pentru a se vindeca de pneumonie).
13. Dacă un agent rațional preferă să obțină recompensa X cu certitudine în locul obținerii recompensei Y cu probabilitatea q, acesta va prefera, de

[1] Există două motive pentru care egalitatea nu poate fi decât aproximativă. În primul rând, șocurile externe vor determina întotdeauna deviații de la egalitate. În al doilea rând, aversiunea față de risc poate determina valori mai mici (și, prin urmare, rate mai mari ale randamentelor) pentru acțiunile puternic volatile.

asemenea, să obțină X cu probabilitatea *p* în loc să obțină Y cu probabilitatea *pq* (axioma independenței din teoria utilității cardinale).

14. Dacă un agent rațional face X când știe că se întâmplă circumstanțele C (sau când intenționează să facă X știind că trebuie să se aștepte la circumstanțele C) și face X în cazul în care circumstanțele C nu se întâmplă (sau intenționează să facă X chiar dacă nu se așteaptă să se întâmple C), atunci ar trebui să intenționeze să facă X, chiar dacă nu știe nimic despre circumstanțe.
15. Un agent rațional nu va face niciodată o ofertă dacă acceptarea acesteia va scoate la iveală informații în urma cărora întreaga afacere va avea o valoare așteptată negativă.
16. Dacă o ofensă determină dorința de răzbunare, atunci persoana ofensată, dacă este rațională, va aștepta până când va putea răspunde cu șanse maxime de succes și șanse minime de risc pentru sine.[1]
17. Un agent rațional provocat la duel va lua lecții de scrimă dacă nu are de ales și trebuie să răspundă.
18. Înainte să ceară cuiva mâna, un agent rațional va culege informații despre dispozițiile comportamentale și emoționale ale celuilalt.

Încălcări ale canonului

Se pare că aceste principii constrângătoare din punct de vedere normativ sunt încălcate adesea. În cele ce urmează am enumerat câteva exemple (cifrele corespund principiilor încălcate).[2]

1. *Paradoxul votului.* Din moment ce niciun scrutin electoral nu a fost câștigat printr-un singur vot, un vot individual nu aduce nicio schimbare

[1] Ar putea exista tentația înlocuirii lui „sau" cu „și" în această propoziție. Însă, exceptând cazurile dominate de întâmplare, nimeni nu își poate maximiza două obiective în același timp. Pentru a fi mai exact, agentul ar căuta dozajul optim fezabil dintre două scopuri, ca fiind reprezentat de un set de oportunitate și o familie de curbe de indiferență (capitolul 9).

[2] Multe dintre exemple au fost citate în capitolul precedent și sunt rezumate aici pentru caracterul lor convenabil. Originea lor este diversă: proverbe, autori clasici, experimente mentale, experimente de laborator și observații din viața de zi cu zi. Exemplele din primele trei categorii se bazează, însă, pe teorii cunoscute, care sunt trecute în revistă în ultima parte a capitolului.

BCU CENTRALA UNIVERSITY LIBRARY

a rezultatului și poate implica un efort considerabil pentru votant. Cu toate acestea, oamenii votează în număr considerabil.[1]

2. *Preferarea unei dureri mai mari decât a uneia mai mici.* Într-un experiment, subiecții au fost expuși la două serii de zgomote foarte neplăcute. Ambele serii au implicat expunerea la 35 de secunde de zgomot de intensitate mare. În cazul primei serii, această expunere a fost urmată de 15 secunde de zgomot care se diminuează gradual (rămânând în continuare neplăcut). Întrebați ce secvență ar prefera să asculte din nou, subiecții au ales-o pe cea mai puțin plăcută în mod neambiguu.
3. *Paradoxul tunderii gazonului de pe peluză.* Într-o comunitate suburbană, domnul H. tunde gazonul de pe peluza din fața casei. Copilul vecinului ar tunde-o pentru doisprezece dolari. Domnul H. nu ar tunde peluza de exact aceeași mărime din fața casei vecinului nici pentru douăzeci de dolari.
4. *Paradoxul clubului de Crăciun.* În acest sistem, clienții depun o sumă lunară cu o dobândă foarte mică sau egală cu zero, sumă pe care o pot retrage doar de Crăciun. Opțiunea de a câștiga o dobândă normală și de a retrage banii când vor fără comision le este, de asemenea, pusă la dispoziție.
5. *Paradoxul cardului de credit.* Când au fost introduse cardurile de credit, grupurile care făceau lobby în favoarea acestora preferau ca orice diferență dintre clienții care plăteau cu bani gheață și cei care foloseau cardul să fie etichetată drept o reducere pentru bani gheață, și nu o suprataxare a cardului. Cu toate că cele două descrieri sunt logic echivalente, clienții utilizau mai degrabă cardul dacă diferența era descrisă ca o reducere în bani gheață.
6. *Erorile jucătorului de ruletă.* Dacă roșul a apărut de mai multe ori la rând, cam jumătate dintre jucătorii de ruletă cred că șansele sunt mai mari de 50% ca la următoarea tură să iasă negrul. Cealaltă jumătate cred că șansele sunt mai mici de 50% să iasă negru.
7. *Eroarea costului investit.* „Să oprești un proiect în care s-au investit 1,1 miliarde de dolari înseamnă să te joci într-un mod iresponsabil cu banii

[1] Paradoxul apare când singurul scop al votanților este acela de a pune un candidat într-o anumită funcție sau de a implementa o propunere. Nu este neapărat nevoie să apară când scopul este acela de a contribui la vitalitatea sistemului democratic sau de a acorda un „mandat" unui candidat pentru că, în aceste cazuri, votul contează chiar dacă nu este esențial. Însă, chiar atunci când motivația este aceea de a sprijini democrația, explicația poate să rezide într-unul dintre mecanismele iraționale care vor fi discutate mai târziu.

plătitorilor de taxe." (Senatorul Denton, 4 noiembrie 1981). Eroarea este numită uneori „eroarea Concorde", după proiectul aeronautic comun între Anglia și Franța, sau „eroarea Vietnam", numit astfel în urma ezitărilor Statelor Unite de a se retrage din misiunea din Vietnam. Dacă ai cumpărat bilete pentru un eveniment, însă o zăpadă grea face imposibilă deplasarea la locul de desfășurare, încă te mai poți duce, chiar dacă, având în vedere actuala situație, ai fi refuzat biletele dacă ți s-ar fi oferit gratis.

8. *Paradoxul dentistului.* Pe 1 martie mă programez la dentist pentru 1 aprilie. Pe 30 martie sun să amân din cauza unei înmormântări (fictive) în familie, la care trebuie să particip. Cu excepția simplei treceri a timpului, nu s-a schimbat nimic în tot acest interval. În special, durerea provocată de măsea este aceeași.
9. *Cele mai bune și cele mai rele scenarii.* Pacienții bolnavi de cancer în stadiu terminal își supraestimează adesea șansele de supraviețuire. În locul unei terapii paliative pentru atenuarea durerii, aceștia aleg chemoterapia agresivă și dureroasă, care are puține beneficii. Întrebați cât de mult ar fi dispuși să plătească pentru a reduce riscul unui dezastru cu probabilitate foarte mică, oamenii sunt dispuși să plătească la fel de mult pentru a reduce riscul la unul într-un milion pe cât ar fi dispuși să plătească pentru a reduce riscul la unul în zece milioane.
10. *Paradoxul bonusului în acțiuni.* Din punct de vedere istoric, beneficiile pe acțiuni sunt mult mai mari decât beneficiile pe obligațiuni. O persoană care investește un dolar în acțiuni pe 1 ianuarie 1928 va avea pe 1 ianuarie 1998 un portofoliu care valorează 1 800 de dolari. Cineva care investește un dolar în obligațiuni va avea un portofoliu care valorează 15 dolari. Problema este de ce această discrepanță nu a dus la o creștere a valorii acțiunilor pentru a apropia beneficiile pe acțiuni de acela pe obligațiuni.
11. *Efectul alternativelor irelevante.* Dacă fiecare dintre două opțiuni A și B este superioară celeilalte din punctul de vedere al uneia dintre două dimensiuni relevante, oamenilor le vine greu să aleagă și se decid să adune mai multe informații despre cele două opțiuni. Dacă introducem o a treia opțiune, C, care este (1) inferioară lui A în privința ambelor dimensiuni și (2) inferioară lui B în privința unei dimensiuni și superioară în privința celeilalte, există o tendință de a alege A fără investigații ulterioare.
12. *Paradoxul apei reci.* Într-un experiment, subiecții cărora li s-a dat de înțeles că durata pe care o pot petrece ținându-și mâinile în apă rece ca gheața este cel mai bun indicator al longevității și-au ținut mâinile în apă mai mult decât aceia cărora nu li s-a dat această informație (falsă).

13. *Efectul certitudinii* (capitolul 7). În experimente, cei mai mulți dintre subiecți preferă să câștige o excursie de o săptămână în Anglia oferită cu certitudine față de șansa de 50% de a câștiga o excursie de trei săptămâni în Anglia, Franța și Italia, însă majoritatea preferă, de asemenea, o șansă de 5% pentru cea de-a doua opțiune în locul unei șanse de 10% pentru prima opțiune.
14. *Efectul de disjuncție.* Dacă subiecții unui experiment se așteaptă să câștige un pariu viitor și sunt întrebați dacă vor fi de acord să ia parte la un pariu ulterior, aceștia au tendința să spună da. Dacă se așteaptă să piardă, sunt dispuși să enunțe aceeași intenție. Însă este mult mai puțin probabil să se pronunțe astfel dacă nu știu fie că vor pierde, fie că vor câștiga. Același efect apare în cazul Dilemei Prizonierului, care presupune o interacțiune unică: este mai probabil ca o persoană să coopereze dacă știe că și celălalt cooperează și – acesta fiind efectul de disjuncție – *chiar mai probabil să coopereze* dacă nu știe nimic despre alegerea celuilalt.
15. *Blestemul învingătorului.* În acest experiment, subiecții au fost rugați să liciteze pentru o parcelă de teren și li s-a spus că vânzătorul știe exact valoarea acesteia, în timp ce ei știu doar că valoarea se încadrează în anumite limite, oricare dintre valorile incluse între aceste limite fiind în egală măsură probabilă. Cumpărătorilor li se spune, de asemenea, că, dacă achiziționează parcela de teren, aceasta va fi cu 50% mai valoroasă pentru ei decât pentru vânzător, deoarece ei vor fi în măsură să o exploateze mult mai eficient. Dacă este acceptată o ofertă, cumpărătorii raționali ar trebui să fie în măsură să infereze *din acest fapt* că valoarea așteptată a terenului pentru ei este mai mică decât valoarea licitată. Dacă valorile se încadrează între 0 și 1 000 și se acceptă suma licitată de (să zicem) 600, cumpărătorul poate infera că valoarea reală a vânzătorului este undeva între 0 și 600, cu o valoare așteptată de 300. Prin urmare, valoarea așteptată pentru cumpărător ar fi 450, adică mai puțin decât se oferă acesta să plătească. Din moment ce argumentul poate fi reprodus pentru orice sumă licitată acceptată, cumpărătorii raționali nu ar trebui să liciteze. Însă, în experimente (inspirate de cazuri reale), toată lumea licitează.
16. *Graba de a căuta răzbunare.* Un proverb spune că „răzbunarea se servește cel mai bine la rece". Un altul spune că „întârzierea în răzbunare face lovitura mai puternică". Probabil ambele proverbe au apărut ca reacție la răzbunarea în pripă, atestând astfel existența acestui fenomen.
17. *Lipsa de considerație pentru eficiență.* Montaigne scrie că „onoarea luptei constă în rivalitatea pasiunilor, nu a experienței; de aceea, i-am văzut pe unii dintre prietenii mei care sunt maeștri ai luptei alegând pentru

duelurile lor arme care îi lipsesc de mijlocul prin care își pot pune în valoare avantajele și care depind în întregime de noroc și dârzenie, astfel încât nimeni să nu poată atribui victoria lor talentului de spadasin, ci curajului".

18. *Cine-n pripă se însoară, toată viața-i e amară.* Această zicală se aplică nu doar căsătoriilor propriu-zise. Când oamenii se îndrăgostesc de o casă sunt uneori atât de dornici să semneze contractul, încât nu sunt atenți la defectele pe care le vor descoperi mult mai târziu.

Alternative la teoria alegerii raționale

Avem acum la dispoziție un set larg de alternative la teoria alegerii raționale pentru a da seamă de aceste probleme. Înainte să le discutăm pe rând, voi enumera mecanismele-cheie ale acestor abordări alternative (punând numărul problemei între paranteze). Unele probleme sunt enumerate de mai multe ori pentru că se poate da seama de ele în mai multe moduri.

- Aversiunea față de pierdere (3, 5, 7, 10)
- Ponderarea neprobabilistă a rezultatelor (13)
- Deprecierea hiperbolică (4, 8)
- Euristica (2, 6)
- Gândirea deziderativă (9, 12)
- Incapacitatea de a proiecta (15)
- Dorința de a acționa dintr-un motiv (11, 14)
- Gândirea magică (1, 12, 14)
- Imperativul categoric (1)
- Emoțiile (3, 7, 14, 18)
- Normele sociale (1, 3, 16, 17)

În gândirea curentă, cele mai remarcabile mecanisme sunt, probabil, aversiunea față de pierdere și deprecierea hiperbolică. După părerea mea, emoțiile sunt o sursă chiar mai importantă de comportament irațional, fie că operează direct sau prin intermediul normelor sociale. Cu toate că emoțiile pot perturba raționalitatea în multe moduri, cel mai important este acela al determinării unei presiuni.

AVERSIUNEA FAȚĂ DE PIERDERE se definește în raport cu un punct de referință, pornind de la presupunerea că oamenii acordă valoare schimbărilor în raport cu un standard de bază mai degrabă decât stărilor finale care rezultă în urma schimbării. Punctul de referință este considerat a fi, în mod obișnuit, statu-quo-ul, deși subiecților le pot fi induse și alte preferințe. Aversiunea față de pierdere reprezintă tendința oamenilor de a acorda o valoare mai mare (în termeni absoluți) unei pierderi situate

la nivelul de referință decât unui câștig de aceleași proporții.[1] S-a descoperit pe cale empirică faptul că raportul este de aproximativ 2,5 la 1. Voi prelua această descoperire în cele ce urmează. O altă proprietate importantă a funcției de valorizare este aceea de a fi concavă pentru câștiguri și convexă pentru pierderi, ceea ce înseamnă că fiecare unitate suplimentară de câștig este valorizată mai puțin decât cea precedentă și fiecare unitate suplimentară pierdută este mai puțin dureroasă decât cea precedentă.

Două dintre probleme pot fi rezolvate prin simplul fapt că pierderile se întrevăd mai mari decât câștigurile. Pentru a rezolva paradoxul cu tunsul peluzei trebuie doar să observăm că aversiunea față de pierdere prezice că banii dați din buzunar și costurile de oportunitate sunt valorizate foarte diferit. Din moment ce valoarea pentru proprietar a unui câștig de 20 de dolari este echivalentă cu valoarea unei pierderi de opt dolari, acesta preferă să renunțe la câștig decât să plătească din buzunarul propriu 12 dolari. Am putea explica prin același raționament și paradoxul referitor la cardurile de credit.

Rezolvarea paradoxului referitor la bonusurile în acțiuni necesită o premisă în plus, aceea că oamenii aleg pachetul mixt de obligațiuni și acțiuni într-un orizont de timp scurt. Din moment ce beneficiul pe acțiuni este volatil, în timp ce obligațiunile produc un venit stabil de la an la an, putem considera că deținerea de acțiuni implică acceptarea unui pariu riscant. Să presupunem că oferim unei persoane un pariu pe acțiuni care îi aduc o șansă de 50% de a câștiga 200 de dolari și o șansă de 50% de a pierde 100 de dolari, având ca punct de referință randamentul fix pe obligațiuni. Dacă luăm aversiunea față de pierdere, putem surprinde acest lucru spunând că valoarea banilor este egală cu x pentru $x > 0$ și egală cu $2{,}5x$ pentru $x < 0$. Din moment ce valoarea unei pierderi de 100 de dolari este egală (în termeni absoluți) cu un câștig de 250 de dolari, perspectiva unui câștig de 200 de dolari nu poate compensa perspectiva, la fel de probabilă, a unei pierderi de 100 de dolari. Prin urmare, persoana va respinge oferta. Să presupune acum că îi oferim un pachet constând în două astfel de pariuri care urmează a fi derulate în perioade succesive. Acest pariu compus este echivalent cu o șansă de 25% de a câștiga 400 de dolari, o șansă de 50% de a câștiga 100 de dolari și o șansă de 25% de a pierde 200 de dolari. Dacă înmulțim pierderea

[1] Presupunând că bunurile (inclusiv banii) au o utilitate marginală descrescătoare, teoria standard a alegerii raționale prezice, de asemenea, că pierderile vor conta mai greu decât câștigurile de dimensiuni comparabile luând același nivel de referință. Magnitudinea efectului, însă, este în mod obișnuit mult mai mică. De asemenea, teoria standard a utilității presupune că plusul de utilitate câștigat înlocuind A cu B este egal cu pierderea de utilitate a înlocuirii lui B cu A, din moment ce aceste diferențe sunt derivate simplu prin compararea *nivelurilor* de utilitate ale celor două stări.

cu 2,5 și o facem comparabilă cu câștigurile, calculând apoi valoarea așteptată[1], aceasta rezultă foarte ușor a fi 25. Prin urmare, persoana va accepta pariul compus. Studiile empirice sugerează că investitorii au tendința de a-și reevalua portofoliile prea des, o practică mioapă care îi determină să investească prea mult în acțiuni și prea puțin în obligațiuni.[2]

Rezolvarea paradoxului referitor la costurile investite face apel doar la forma funcției de valorizare (concavitatea sau convexitatea). Să luăm în considerare următorul exemplu. O familie plătește p dolari pentru bilete la un meci care urmează să se desfășoare la 96 de kilometri distanță. În ziua meciului se produce o furtună de zăpadă. Cu toate acestea, se hotărăsc să meargă, dar observă în treacăt că, dacă li s-ar fi dat biletele gratis, ar fi preferat să stea acasă. Notând cu v funcția de valorizare pentru câștiguri, valoarea mersului la meci este $v(g)$. Notând cu v^* funcția de valorizare pentru pierderi, valoarea pierderii a p dolari este numărul negativ $v^*(-p)$. Costul suportării furtunii de zăpadă este c. Presupunem că $v(g) = -v^*(-c)$, ceea ce implică faptul că, dacă membrii familiei ar fi primit biletele gratis, atunci aceștia ar fi fost indiferenți între a sta acasă și a se duce la meci prin furtună. Dar, din moment ce aceștia au plătit deja p dolari, preferă să meargă. Pentru a pune acest lucru în evidență, trebuie să observăm că, din cauza convexității lui v^*, $v^*(-(-c + p)) > v^*(-c) + v^*(-p)$.[3] Putem scrie ultima relație ca $v^*(-(c + p)) - v^*(-c) > v^*(-p)$, care este echivalentă, pe baza presupoziției enunțate, cu $v^*(-(c + p)) + v(g) > v^*(-p)$. Din moment ce termenul din stânga al ultimei inegalități este expresia netă a câștigului sau a pierderii produse de mersul la meci, iar termenul din dreapta este pierderea produsă de faptul de a nu merge, aceștia preferă să meargă.

PONDERAREA NEPROBABILISTĂ A REZULTATELOR. Aversiunea față de pierdere decurge dintr-o alternativă influentă la teoria alegerii raționale, numită *teoria perspectivei*. O altă implicație a acestei teorii este că oamenii tind să ponderеze

---

[1] Din moment ce teoria perspectivei presupune că ponderile deciziilor sunt diferite de probabilități, acest calcul este doar aproximativ corect. Nici această simplificare, nici asumpția unei funcții de valorizare liniare nu contează pentru concluzia acestei analize.

[2] Termenul „miop" nu are același înțeles aici cu acela pe care l-a avut în analizele legate de deprecierea timpului (capitolul 6). Nu se referă la modul în care agentul calculează valoarea prezentă a unor fluxuri viitoare de venituri, ci la tendința de a lua decizii succesive în mod separat, și nu de a le trata „la pachet" într-o decizie globală. O astfel de „miopie a deciziei", cum am putea să o numim, ar putea funcționa și în alte contexte. Astfel, când oamenii încearcă să controleze deprecierea hiperbolică prin „înmânuncherea" unor alegeri succesive (capitolul 13), succesul acestora poate depinde de numărul de alegeri inclus.

[3] Din moment ce este o comparație a două numere negative, inegalitatea spune că primul termen este mai aproape de zero, adică este mai mic ca valoare absolută.

rezultatele în mod diferit de ceea ce susține teoria utilității așteptate. Conform acestei teorii, utilitatea este liniară în probabilități (capitolul 11). Dimpotrivă, teoria perspectivei susține că oamenii sunt mai sensibili la schimbările de probabilitate în zona limitelor naturale de 0 (imposibil) și 1 (cert). Efectul certitudinii pune în evidență neliniaritatea în jurul lui 1. Creatorii teoriei perspectivei, Daniel Kahneman și Amos Tversky, dau următorul exemplu (pe care îl atribuie lui Richard Zeckhauser) al neliniarității în jurul lui 0:

> Să presupunem că ești constrâns să joci ruleta rusească, dar ți se dă șansa de a cumpăra eliminarea unui glonț din pistolul încărcat. Ai plăti la fel de mult ca să reduci numărul de gloanțe de la patru la trei pe cât ai plăti ca să reduci numărul lor de la unu la zero? Cei mai mulți oameni simt că ar fi dispuși să plătească mult mai mult pentru o reducere a șansei de moarte de la 1/6 la zero decât pentru o reducere de la 4/6 la 3/6. Considerațiile economice [adică teoria utilității așteptate] ne-ar duce la concluzia că ar trebui să plătim mai mult în ultimul caz, în care valoarea banilor este, fără îndoială, redusă prin probabilitatea considerabilă a faptului că nu vei apuca să te bucuri de ea.

DEPRECIEREA HIPERBOLICĂ a fost discutată în capitolul 6. Aici vreau doar să adaug observația că există o legătură foarte strânsă între problemele (8) și (4). Motivul pentru care oamenii iau parte la cluburile de Crăciun este acela că știu că, dacă își pun economiile într-un cont normal cu intenția de a le ține acolo până la Crăciun, se vor răzgândi și le vor retrage.

EURISTICA. Euristica (sau regula empirică) îi poate induce pe oameni în eroare. Credința pariorului că roata ruletei are memorie poate veni fie din euristica reprezentativității („Ar fi timpul să apară roșu"), fie din euristica disponibilității („Roșu va avea succes acum"). Preferința pentru zgomotul mai neplăcut își are originea în utilizarea unei euristici de tip „vârf – final", conform căreia experiențele trecute sunt valorizate în funcție de cum au fost în punctul lor de vârf (pozitiv sau negativ) și cum au fost la final, iar nu prin caracterul lor plăcut sau neplăcut per ansamblu. O astfel de euristică ar determina oamenii să prefere experiența care este cea mai rea din punct de vedere obiectiv, din moment ce are un final mai bun decât cealaltă și același vârf.

GÂNDIREA DEZIDERATIVĂ. Fenomenul gândirii dezideretive a fost discutat în capitolul 7. Aceasta poate fi declanșată de o simplă dorință, ca în cazul oamenilor care au locuri de muncă riscante și bine plătite, dar care diminuează riscurile cu care se confruntă. Este cu atât mai probabil să se manifeste când dorința își are originea într-o emoție puternică, așa cum se întâmplă în cazul pacienților bolnavi de cancer care aleg tratamentul al cărui efect este, în cele din urmă, acela de a-i face să sufere mai mult.

Gândirea deziderativă determinată de dorința pentru o viață lungă poate fi responsabilă și pentru paradoxul apei reci.

INABILITATEA DE A PROIECTA. Într-un număr de situații, oamenii iau decizii proaste pentru că nu sunt capabili să se proiecteze în viitor. Vreau să spun prin aceasta că le lipsește capacitatea de a-și imagina ce motive ar avea să creadă (ei sau ceilalți) sau ce anume i-ar îndemna să acționeze în situații viitoare care depind de alegerea prezentă. Blestemul învingătorului poate fi explicat prin această incapacitate. Pentru a da un alt exemplu, putem lua în considerare cazul președintelui Chirac, a cărui cerere pentru alegeri anticipate, în iunie 1997, a avut consecințe dezastruoase pentru el însuși. Motivul pentru care coaliția sa a pierdut rezidă, poate, în faptul că votanții au înțeles că, dacă președintele vrea alegeri mai devreme, atunci acest lucru se întâmplă pentru că știe ceva ce ei nu știu și care îl face să creadă că ar pierde dacă așteaptă. Solicitând alegeri anticipate, președintele a devoalat ce știa sau, cel puțin, a devoalat *faptul că* știa ceva nefavorabil și le-a dat votanților un motiv să voteze împotriva sa. Sondajele i-au spus că va câștiga, dar *sondajele nu seamănă cu alegerile*, din moment ce efectuarea unui sondaj nu le dezvăluie nimic respondenților cu privire la credințele persoanei care a cerut respectivul sondaj.[1]

DORINȚA DE A ACȚIONA DINTR-UN MOTIV. Am menționat mai multe instanțe ale acestui mecanism la începutul capitolului de față. În paradoxul 11, dorința îl determină pe agent să își modifice comportamentul odată cu introducerea unei opțiuni care este în mod neambiguu inferioară uneia dintre opțiunile deja disponibile.[2] În alte cazuri, adăugarea unor opțiuni îl poate împiedica pe agent de la a lua orice decizie. Au fost întâlnite și în viața de zi cu zi comportamente similare cu acela al măgarului lui Buridan, care a murit de foame pentru că nu a putut alege dintre două grămezi identice de fân. Un psiholog a instalat pe Broadway mai multe standuri la care vindea gem și a descoperit că, atunci când standul prezenta o varietate mare de mărci, trecătorii se uitau la mai multe, dar cumpărau mai puțin în comparație cu standurile care prezentau o varietate mai mică. Cu mai multe opțiuni la dispoziție este mai greu să-ți

---

[1] În Partea a V-a discut comportamentul lui Chirac într-o situație care s-ar putea numi „sindromul fratelui mai mic".

[2] Exprimarea ar putea induce în eroare întrucât, în experimente, nu sunt *aceiași* subiecții expuși primului set de alegeri (A, B), apoi setului (A, B, C), ci două grupuri diferite de subiecți distribuiți aleatoriu. Interpretarea firească a descoperirii, însă, este aceea că subiecții din grupul (A, B, C) *s-ar fi comportat* precum subiecții din grupul (A, B) dacă ar fi fost expuși doar la cele două opțiuni. Prin urmare, „modificarea" se referă la un punct de referință contrafactual, nu la unul real. Această observație se aplică multor experimente menționate în carte.

spui fără ezitare „Acesta e cel mai bun produs". Cei care trebuie să își bazeze alegerea pe rațiune suficientă nu vor alege nimic.

După cum sugerează paradoxul 14, pentru a acționa dintr-un motiv trebuie să *fii în posesia* unui motiv, și nu doar să *știi* că ești în posesia unui motiv. Astfel, să presupunem că știu că trebuie să fie una dintre alternativele *p* și *q*, dar nu știu care anume. Dacă *p* este alternativa corectă, am un motiv să fac X. Dacă *q* este alternativa corectă, am, de asemenea, un motiv să fac X. Prin urmare, știu că, indiferent de caz, am un motiv să fac X, dar, din moment ce nu știu *care* este motivul, mă abțin de la a face X. Problema 14 oferă un exemplu al acestei anomalii.[1] Pentru alt exemplu, putem lua în considerare faptul că, în vechiul cod juridic englez, un acuzat putea fi achitat în cazul în care probele nu clarificau dacă infracțiunea comisă era furtul sau delapidarea, cu toate că inculpatul ar fi fost condamnat dacă s-ar fi dovedit oricare dintre acuzații. Pentru a fi condamnat ar fi trebuit sau să fie găsit vinovat de *p* sau de *q*. Faptul de a fi fost găsit vinovat de (*p sau q*) nu ar fi fost suficient. Un alt răspuns pentru a ieși din încurcătură ar fi amânarea luării unei decizii. Să presupunem că guvernul Statelor Unite dezbate dacă să stabilească o relație comercială cu un stat în care urmează să se desfășoare alegeri prezidențiale. În cazul în care câștigă candidatul proamerican, Statele Unite vor avea o rațiune suficientă pentru a stabili relația comercială. În cazul în care câștigă candidatul antiamerican, Statele Unite vor avea un alt set de rațiuni suficiente pentru a urma aceeași politică. Însă Statele Unite ar putea să aștepte până după alegeri pentru a-și anunța politica. În anumite circumstanțe, aceasta ar putea fi o amânare costisitoare și de prisos.

GÂNDIREA MAGICĂ. Mecanismul gândirii magice (capitolul 7) ar putea explica comportamentul în cazul paradoxului apei reci. Ar putea explica, de asemenea, unele dintre cazurile în care întâlnim efectul de disjuncție. Dacă este mai probabil ca oamenii să coopereze în Dilema Prizonierului când aceștia nu știu dacă cealaltă persoană este cooperantă sau nu, poate fi pentru că aceștia cred, în chip magic, că prin cooperarea lor îl pot face pe celălalt să coopereze. „Fiind ca mine, va acționa ca mine." Și intențiile de vot pot fi modelate astfel. Cred în mod irațional că votul meu nu este doar un element predictiv al votului altora, ci, cumva, că face mai probabilă prezența la vot a celorlalți; eficiența crescută a acțiunii mele face votul să pară rațional.

IMPERATIVUL CATEGORIC. Există o relație strânsă între această ultimă instanță a gândirii magice, imperativul categoric, și versiunea simplificată a imperativului

[1] Într-o manieră asemănătoare, suma maximă pe care oamenii sunt dispuși să o plătească pentru o loterie între două opțiuni s-a demonstrat a fi mai mică decât suma maximă pe care sunt dispuși să o plătească pentru cea mai puțin atractivă dintre opțiuni.

categoric kantian, conform căreia o persoană ar trebui să facă A mai degrabă decât B pentru că ne-ar fi tuturor mai bine dacă am face A decât dacă am face B. Faptul de a acționa pe baza imperativului categoric este, însă, irațional. Raționalitatea îmi spune să aleg în funcție de ceea ce se va întâmpla dacă *eu* fac A mai degrabă decât B.[1] Imperativul categoric îmi spune să aleg în funcție de ce se va întâmpla dacă toată lumea face A mai degrabă decât B. În cadrul alegerilor naționale, chiar și aceia care nu sunt subiecți ai gândirii magice se pot „abține de la a se abține" de la gândul „Oare ce-ar fi dacă toată lumea ar face la fel?".

EMOȚIILE. Pentru a compara comportamentul bazat pe emoții cu comportamentul rațional, putem modifica Figura 11.1 astfel încât aceasta să includă (săgețile îngroșate) impactul emoțiilor asupra fiecărui element al Figurii (vezi Figura 12.1).

S-a susținut că emoțiile pot afecta *acțiunea* în mod direct, în cazul slăbiciunii voinței (capitolul 6). Medeea, ucigându-și copiii pentru a se răzbuna pe Iason, știe că acționează împotriva propriei judecăți despre ce este bine. Mi-am expus deja rezervele față de această idee, dar ea nu poate fi exclusă. Emoțiile afectează *dorințele* în două moduri. În primul rând, prin tendința asociată acțiunii, acestea pot cauza o modificare temporară a preferințelor. Dacă situația impune o întârziere între momentul la care se ia decizia de a acționa și momentul acțiunii, atunci este posibil ca acțiunea să nu mai aibă loc. Un exemplu menționat în capitolul 8 a fost modul în care o manifestare crescută a interesului pentru înrolarea în armată după atacurile din 11 septembrie 2001 nu a dus la o creștere a recrutărilor. Dacă măsura poate fi pusă în practică

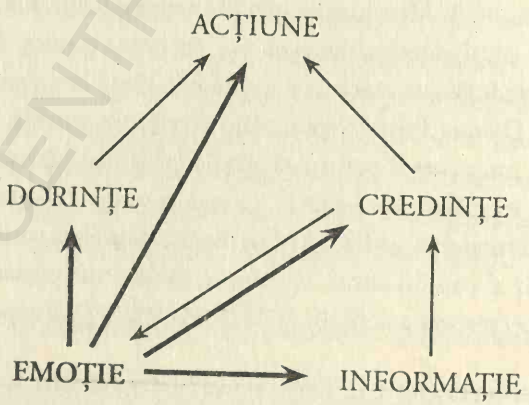

FIGURA 12.1

[1] Raționalitatea nu îmi spune să aleg în funcție de ce mi se va întâmpla dacă fac A mai degrabă decât B (vezi capitolul 11). Este compatibilă cu anumite forme de moralitate care țin cont de celălalt, dar nu cu o formă reprezentată de imperativul categoric.

imediat, ea poate fi ulterior inversată, odată cu calmarea emoției. Astfel, dintre cei două sute de mii de oameni care au dezertat, cel mai probabil de frică, din armata Uniunii în timpul Războiului Civil din Statele Unite, 10% s-au întors voluntar. Însă, în unele cazuri, acțiunea întreprinsă pe baza unor preferințe temporare determinate de emoție este reversibilă (o situație de tip „capcană de homar"[1]). Când tineri bărbați și femei se alătură unei mișcări de gherilă într-un moment de entuziasm, iar mai târziu vor să plece, s-ar putea să descopere că nu au la dispoziție această opțiune. În al doilea rând, emoțiile pot determina o creștere temporară a ratei de depreciere a timpului, astfel încât opțiunile mai puțin preferate până atunci, cu consecințe nefaste pe termen lung, par acum preferabile.

Emoțiile pot afecta *credințele* direct prin mecanismele gândirii deziderative și ale gândirii antideziderative. Emoția aroganței – bazată pe credința în superioritatea propriei persoane – va rezista conștientizării unei greșeli personale. Putem astfel explica, cel puțin în anumite cazuri, vulnerabilitatea la eroarea costurilor investite.[2] Aroganța poate explica și de ce oamenii au tendința de a da vina pe victimele lor[3]: din moment ce imaginea de sine nu le permite să admită vina, aceștia găsesc un defect în celălalt, care nu doar justifică ce le-a făcut acesta, ci motivează și comportamentele dăunătoare ulterioare. La fel ca în cazul erorii costului investit, caracterul refractar în a admite o greșeală din trecut poate conduce la consecințe mai rele decât greșeala însăși. Cât privește emoțiile și gândirea antideziderativă, vom urmări în capitolul 23 câteva exemple de panică determinată de frică.

În cele din urmă, în virtutea presiunii exercitate de emoții (capitolul 8), acestea pot interfera cu dobândirea optimă de *informație* și, prin urmare, pot afecta credințele în mod indirect. Mânia (paradoxul 16) și dragostea (18) ne determină să facem lucruri pe care nu le-am fi făcut dacă am fi urmat o politică mai rațională de culegere de informații. O observație a lui Seneca ne poate ajuta să adunăm toate aceste elemente într-o singură idee: „Mânia este în întregime lipsită de echilibru; acum o ia înainte mai mult decât ar trebui, acum se oprește mai repede decât ar fi cazul." De foarte multe ori, presiunea determină neglijarea efectelor îndepărtate din punct de vedere

---

[1] Capcanele de homari sunt construite în așa fel încât permit capturarea homarilor mari și le permit celor de mici dimensiuni să iasă. (N. trad.)

[2] În acest context, este interesant că animalele nu par a comite eroarea costurilor investite: poate pentru că nu au vreo imagine de sine de care să le pese.

[3] După cum voi sublinia la sfârșitul capitolului 15, există o tendință de a da vina pe victimă pentru o acțiune comisă de un terț, ca atunci când unii oameni, alții decât violatorul, dau vina pe victimă pentru viol. Aici mecanismul nu mai este aroganța, ci o credință sau o presupoziție tacită că, din moment ce lumea este fundamental dreaptă, victima trebuie cumva să-și fi meritat soarta.

temporal ale opțiunilor, în virtutea faptului că *recunoașterea unor consecințe pe termen lung este în sine o chestiune cronofagă*. Prin urmare, trunchierea orizontului de timp nu se datorează neapărat unei deprecieri mai mari a consecințelor cunoscute, ci faptului că unele consecințe nici măcar nu apar pe radarul mental al agentului.

NORMELE SOCIALE. Emoțiile disprețului și rușinii joacă un rol important în consolidarea normelor sociale (capitolul 21). Graba pentru răzbunare poate fi datorată presiunii exercitate de mânie, dar ar putea să se datoreze și unei norme sociale care îl etichetează drept laș pe acela care amână răzbunarea. Refuzul dueliștilor de a alege arma pe care o cunosc cel mai bine este susținută de frica de a fi perceput drept excesiv de preocupat cu propria supraviețuire, și nu de onoare. În cele din urmă, paradoxul referitor la tunsul ierbii poate fi explicat prin acțiunea unor norme sociale, mai degrabă decât prin aversiunea față de pierdere. Un locuitor din cartier nu se poate gândi să tundă peluza vecinului său pentru că există o normă socială în comunitățile suburbane împotriva adulților care se ocupă de astfel de sarcini pentru bani. Pur și simplu, așa ceva *nu se face*. Și votul poate reflecta prezența unei norme sociale dacă actul votului este vizibil altora, iar aceștia au o părere foarte proastă despre aceia care nu votează. Pare mai plauzibil să ne gândim la votul în alegeri anonime de anvergură ca la un rezultat al unor norme *morale* care pot avea și ele aspecte iraționale (vezi discuția de mai devreme).

◈◈◈

Notă bibliografică

O evaluare critică a teoriei alegerii raționale dintr-o perspectivă oarecum diferită de cea oferită aici poate fi găsită în D. Green și I. Shapiro, *Pathologies of Rational Choice Theory* (New Haven, CT, Yale University Press, 1994), adăugită într-un mod foarte util de J. Friedman (ed.), *The Rational Choice Controversy* (New Haven, CT, Yale University Press, 1996). Pentru ideea de raționalitate ca normă, vezi D. Føllesdal, „The status of rationality assumptions in interpretation and in the explanation of action", în M. Martin și L. McIntyre (editori), *Readings in the Philosophy of the Social Sciences* (Cambridge, MA, MIT Press, 1994). Am susținut teza hiperraționalității în cazul principiului „celor mai bune interese ale copilului" în capitolul 3 din *Solomonic Judgments* (Cambridge University Press, 1989). O discuție utilă a unor chestiuni conexe se regăsește în J. Wiener, „Managing the iatrogenic risks of risk management", *Risk: Health, Safety and Environment*, 9 (1998), pp. 39-82. Cele mai multe paradoxuri au fost discutate în articole retipărite din cărțile-sursă enumerate în Nota bibliografică a capitolului 7. Excepție fac paradoxul 1, pentru care vezi A. Blais, *To Vote or Not to Vote:*

The Merits and Limits of Rational Choice Theory (Pittsburgh, University of Pittsburgh Press, 2000), și paradoxul 2, pentru care se poate consulta D. Kahneman, „Objective happiness", în D. Kahneman, E. Diener și N. Schwartz (editori), *Well-Being: The Foundations of Hedonic Psychology* (New York, Russell Sage, 1999); pentru paradoxurile 16 și 17 se poate vedea capitolul 3 din cartea mea *Alchemies of the Mind* (Cambridge University Press, 1999), iar pentru paradoxul 18 se poate urmări discuția din capitolul 8. Cele mai multe dintre alternative sunt trecute în revistă în aceleași cărți-sursă, cu excepția gândirii deziderative (capitolul 7), a emoțiilor (capitolul 8) și a normelor sociale (capitolul 21). Ar trebui să se observe că efectul certitudinii este foarte înrudit cu primul paradox, care a fost prezentat explicit (în 1953) ca o provocare adresată teoriei alegerii raționale, și anume „paradoxul Allais". „Paradoxul Chirac" este dezbătut, împreună cu alte exemple asemănătoare, în A. Smith, *Election Timing* (Cambridge University Press, 2004). Teza conform căreia animalele nu comit eroarea costurilor investite este susținută în H. Arkes și P. Ayton, „The sunk cost and Concorde effects: Are humans less rational than lower animals?", *Psychological Bulletin*, 125 (1999), pp. 591-600. Exemplul cu vânzarea de gem pe Broadway este din S. Iyengar și M. Lepper, „When choice is demotivating: Can one desire too much of a good thing?", *Journal of Personality and Social Psychology*, 79 (2000), pp. 995-1006. Referința la delapidare și furt în vechiul cod juridic englez este din J.F. Stephen, *A History of English Criminal Law* (Londra, Macmillan, 1883; Buffalo, NY, Hein, 1964), vol. 3, p. 153. Descoperirile cu privire la reînregimentarea soldaților armatei unioniste este din D. Costa și M. Kahn, „Deserters, social norms and migration", în curs de apariție în *Journal of Law and Economics*. Dovezi ale unei deprecieri crescute a timpului determinată de emoție, sunt prezentate în D. Tice, E. Braslavsky și R. Baumeister, „Emotional distress regulation takes precedence over impulse control", *Journal of Personality and Social Psychology*, 80 (2001), pp. 53-67.

BCU IASI/CENTRU VECHI

Capitolul 13

REACȚII LA IRAȚIONALITATE

Raționalitate de rang secund

În ultimele două capitole am luat în considerare idealul comportamentului rațional și devierile frecvente de la raționalitate. Aceste devieri, oricât de răspândite și frecvente ar fi, nu sunt inevitabile. Dacă ne vom înțelege înclinația de a face greșeli, vom putea lua – și chiar luăm – măsuri de precauție pentru a reduce șansele de a repeta greșelile sau, cel puțin, pentru a limita răul pe care îl facem. După cum am spus de mai multe ori, *vrem* să fim raționali. Ne putem gândi la aceste strategii de precauție ca la o formă imperfectă de raționalitate sau o raționalitate de rang secund. Ar trebui să le distingem de simpla *învățare*, care apare de fiecare dată când înclinația dispare în urma unei intuiții îmbunătățite. S-a descoperit, de pildă, că în momentul în care oamenii își dau seama că votul este, într-un anumit sens, inutil, șansa ca aceștia să mai voteze scade.[1] Erorile cognitive, care sunt asemănătoare iluziilor optice, pot fi corectate prin învățare. La fel cum învățăm să ignorăm aparența unui băț care pare frânt în apă, unii jucători de ruletă învață, după câte se pare, pe propria piele că zarurile nu au memorie. Aici, însă, mă preocupă înclinațiile care persistă în timp.

Pentru a face față tendințelor noastre de a ne comporta irațional, putem folosi fie strategii intrapsihice, fie mijloace extrapsihice (preangajamente). Mai întâi voi ilustra modul în care sunt folosite aceste tehnici pentru a contracara deprecierea hiperbolică și comportamentul inconsistent pe care aceasta îl generează, iar apoi voi discuta utilitatea lor în controlul emoțiilor și al comportamentelor de dependență. Aceste strategii nu sunt neapărat raționale, dar multe dintre ele sunt.

Sinele viitor ca aliat

Un agent care este subiect al deprecierii hiperbolice și știe aceasta este *sofisticat*. Spre deosebire de agentul naiv, care descoperă că se răzgândește neîncetat fără să

[1] În special studenții la economie par a se comporta astfel.

BUCHAREST UNIVERSITY LIBRARY

înțeleagă mecanismul subiacent, agentul sofisticat este, în același timp, conștient de înclinația sa și o găsește deplorabilă. Anticipând situații viitoare în care va avea de ales între o recompensă mai mică obținută mai devreme și una mai mare obținută mai târziu, agentul ar vrea să se constrângă să o accepte pe cea din urmă, în ciuda înclinației sale de a o alege pe cea dintâi. În anumite cazuri, agentul poate să își trateze „sinele viitor" ca pe un aliat în efortul comun de a înfrânge tentațiile. În alte cazuri, agentul își poate trata sinele viitor ca pe un inamic și poate încerca să limiteze răul pe care acesta îl poate comite asupra „sinelui curent". Firește, acest limbaj este metaforic, dar urmează să îl purificăm de metafore cât de curând.

Să luăm mai întâi în considerare cazul în care alegerea între o recompensă mai mică mai devreme și una mai mare mai târziu poate fi așteptată să apară des. Agentul poate să se constrângă să urmărească recompensa mai târzie prin *gruparea* (sau *împachetarea*) alegerilor.

Voi ilustra printr-un exemplu din vremea când locuiam pe colină, aproape de universitatea unde predam. În fiecare zi mergeam cu bicicleta spre campus și înapoi. Drumul spre casă presupunea cățărări destul de abrupte, așa că în fiecare zi eram pus în fața tentației de a coborî de pe bicicletă și de a merge pe jos, în loc să mă forțez să pedalez. Când ieșeam din campus, îmi promiteam că voi sta pe bicicletă tot drumul, dar, în mijlocul cățărării, mă păștea adesea un gând foarte seducător: „De ce să nu merg pe jos azi și să reiau mersul pe bicicletă de mâine?" Însă, imediat îmi apărea, din fericire, alt gând: „Ce e atât de special la ziua de mâine? Dacă cedez tentației azi, oare acest lucru nu prezice că voi face la fel și mâine, și în ziua de după și așa mai departe?" Ultimul gând îmi îngăduia să rămân pe bicicletă.

Acest instrument intrapsihic presupune o *reîncadrare* a situației. În loc să mă gândesc la drumurile spre casă din viitor ca implicând o serie de alegeri, am început să le văd ca pe o alegere între două serii: să merg întotdeauna pe bicicletă și să merg întotdeauna pe jos. Spunându-mi că, într-o anumită ocazie, comportamentul respectiv este cel mai bun element predictiv pentru o ocazie ulterioară, am pornit un efect de domino intern prin care am ridicat mizele și m-am constrâns să urmăresc recompensa mai târzie a unei sănătăți îmbunătățite decât recompensa mai timpurie a eliberării de disconfort. Putem arăta, cu referire la Figura 6.1, că, de fapt, dacă situăm multe perechi de recompense identice cu A și B pe linia orizontală, iar apoi trasăm două curbe, una pentru suma curbelor valorilor prezente ale tuturor recompenselor mici și alta pentru suma curbelor valorilor prezente ale tuturor recompenselor mari, atunci ultima curbă va fi deasupra celei dintâi în momentul în care trebuie făcută prima alegere, cu condiția ca numărul de alegeri succesive care urmează a fi făcute să fie suficient de mare.[1]

[1] Pentru a ilustra acest lucru, să presupunem că valoarea prezentă a unei unități de utilitate a momentul t în viitor este $1 / (1 + t)$ și că agentul este expus de două ori alegerii între o

Cu alte cuvinte, datorită grupării alegerilor, opțiunea de a urmări întotdeauna recompensa mai mare devine preferabilă urmăririi constante a recompensei mai mici. Sigur, a alege recompensa mai mică azi și recompensa mai mare în toate ocaziile ulterioare e chiar mai bine, dar, prin ipoteză, această opțiune nu se află în setul de oportunitate al agentului.[1]

Poate fi justificată această presupoziție? *Este* oare comportamentul meu de azi un element predictiv bun al comportamentului meu de mâine? În cazuri care implică un efect cazual autentic, acest lucru poate fi adevărat. Mergând cu bicicleta azi, mușchii mei vor rămâne puternici, astfel încât să pot merge cu bicicleta și mâine.[2] Totuși, în cazul meu, m-am bazat mai degrabă pe gândire magică decât pe eficiență cauzală. La fel cum mulți oameni votează sau fac donații sub influența gândului „Dacă nu eu, atunci cine?", am fost ținut pe bicicletă de gândul „Dacă nu acum, atunci când?". Sau, chiar mai elaborat: „Nu e nimic special cu privire la ziua de azi. Dacă mă dau jos de pe bicicletă, cauzele care mă determină să fac acest lucru vor funcționa și mâine și vor induce același comportament. Dacă nu fac acum un efort, nu îl voi face niciodată." În absența unui efect cauzal autentic, însă, concluzia nu decurge. Dacă pot sta pe bicicletă azi, dar decid să cobor, atunci pot sta pe bicicletă și mâine. Cu toate că este fals, raționamentul este constrângător și, cred, foarte răspândit. Ne arată că putem implica

---

recompensă mică de 3 și o recompensă mare de 10. Recompensele mici devin disponibile la momentele 0 și 6, iar recompensele mari la momentele 3 și 9. La momentul 0, valoarea prezentă a primei recompense mari este de 10 / (1 + 3) = 2,5, care este o valoare mai mică decât valoarea prezentă (instantanee) a recompensei mici. Dacă alegerea se face strict pe baza acestei comparații, atunci va fi aleasă recompensa mai mică. Aceeași alegere, din motive identice, va fi făcută la momentul 6. Pentru că suma valorilor prezente ale celor două recompense mici este 3 + 3 / (1 +6) ≈ 3,43 și suma valorilor prezente ale celor două recompense mari este 10 / (1 + 3) + 10 / (1 + 9) = 3,5, atunci gruparea îl va face pe agent să prefere cele două recompense mari.

[1] Dacă agentul suferă de „miopie a deciziei" (capitolul 12), gruparea poate să nu funcționeze. Să presupunem, de pildă, că agentul grupează alegerile cu mult înainte de momentul la care devine disponibilă prima recompensă mică. La momentul grupării, valoarea prezentă a fluxului de recompense mari este mai mare decât aceea a fluxului de recompense mici, iar agentul intenționează în mod ferm să aștepte prima recompensă mare. Pe măsură ce se apropie în timp de momentul când prima recompensă devine disponibilă, intenția poate sau nu să supraviețuiască. Această inversare a preferințelor nu se datorează deprecierii hiperbolice în sine, ci miopiei deciziei.

[2] Pentru a ilustra cu o analogie, să presupunem că, votând, îi pot influența pe mulți alții care, altfel, nu ar fi prezenți la scrutin. Observați că, pornind de la această presupoziție, nu este implicată gândirea magică, ci doar un efect cauzal multiplicativ.

un tip de iraționalitate (gândirea magică) pentru a combate un alt tip (deprecierea hiperbolică).[1]

Pentru o bună funcționare, poate că ar trebui să construim aceste strategii ca alegeri binare: ori o faci întotdeauna, ori niciodată. Pentru mulți oameni, abținerea este mai ușoară decât moderația. Boswell observă că „Johnson, cu toate că putea fi *abstinent*, nu era un om *temperat* nici la mâncare, nici la băutură. Putea să se abțină, dar nu putea să consume moderat." Ne confruntăm cu aceeași problemă dacă, în loc să limităm consumul cu fiecare ocazie, încercăm să limităm numărul de ocazii în care ni-l putem permite. Stratagema de a decide dinainte ce înseamnă o ocazie în care ne putem permite anumite lucruri se erodează foarte repede. Când oamenii se decid să nu mai bea alcool înainte de cină, se pot găsi în situația de a fixa ora cinei din ce în ce mai devreme. Regula de a consuma vin doar la restaurant, niciodată acasă, poate determina o persoană să ia masa mai des în oraș. Regula lui Kant de a fuma doar o singură pipă după micul dejun (capitolul 4) nu era suficient de neambiguă pentru a-i oferi protecție totală din moment ce, odată cu trecerea timpului, acesta și-a cumpărat pipe din ce în ce mai mari. Regula „Nu face asta niciodată" poate fi singura care poate fi urmată consecvent, atunci când este fezabilă. Din moment ce o astfel de politică nu este fezabilă cu privire la mâncare și băutură, obezitatea poate fi mai recalcitrantă decât sunt dependențele în raport cu regulile private.

Construcția binară poate determina însă un comportament absurd de rigid. Să presupunem că mi-am propus să nu fac nicio excepție de la regula de a mă spăla pe dinți în fiecare seară. Într-o anumită ocazie mă pot trezi fără periuță de dinți la îndemână și pot decide să merg opt kilometri prin furtună pentru a-mi cumpăra una. Pentru a întări decizia, îmi spun că, dacă încalc regula cu acea ocazie, voi cădea pe o pantă alunecoasă care mă va face să încalc regula și din motive mai banale, iar apoi regula va înceta să funcționeze și-mi vor cădea dinții. Unii oameni își construiesc sisteme foarte elaborate de acest tip, în care eșecul de a urma o regulă este un element predictiv al eșecului cu privire la alte reguli, în felul acesta miza mărindu-se din ce în ce mai mult.[2] Întrucât regulile private pot avea aceste efecte absurde, de multe ori ele oferă un remediu mai periculos decât boala. În limbaj freudian (capitolul 4), controlul rigid al pulsiunilor, exercitat de supraeu, poate face mai mult rău decât pulsiunile sinelui.

---

[1] În cazul votului, efectul gândirii magice este acela de a ne ajuta să învingem înclinația spre un comportament rațional cu potențial dăunător pentru societate, mai degrabă decât aceea de a contracara iraționalitatea.

[2] Aceștia pot crede în mod greșit (capitolul 10) că existența consecvenței de la o situație la alta va determina declanșarea unor colapsuri de la o situație la alta.

Sinele viitor ca inamic

Să luăm acum în considerare cazul în care agentul se confruntă cu alegerea dintre recompense (sau pedepse) într-un anumit moment viitor. (Spre deosebire de cazul precedent, se presupune că alegerea apare o singură dată.) Agentul poate să adopte instrumentul intrapsihic de a răspunde strategic înclinației cunoscute a „sinelui viitor" de a deprecia viitorul hiperbolic. Să presupunem că sunt genul de om care amână hiperbolic lucrurile, lăsând pe mâine, apoi pe poimâine etc. Odată ce înțeleg că sunt supus unei astfel de înclinații, comportamentul meu optimal se modifică. Să presupunem că pot duce la îndeplinire o sarcină foarte neplăcută în oricare dintre trei perioade date, costul crescând cu trecerea timpului. Dacă sunt naiv, îmi pot spune că voi rezolva sarcina mâine. Dacă sunt sofisticat, știu că mâine voi amâna până în ultima perioadă de timp disponibilă. Înțelegerea faptului că mă va aștepta un cost foarte mare dacă nu rezolv imediat sarcina mă poate determina să fac tocmai asta.[1]

În acest caz, este bine să fii sofisticat. În alte cazuri, a fi naiv e și mai bine. Să presupunem că poți obține o recompensă în oricare dintre trei perioade succesive și că recompensa crește cu timpul. Un exemplu ar fi o persoană căreia i se oferă o sticlă de vin care se îmbunătățește cu trecerea timpului, până în al treilea an, când începe să se degradeze. O persoană naivă își poate forma intenția să aștepte până în perioada a treia, dar, pe urmă, se poate răzgândi și va bea vinul în perioada a doua. O persoană sofisticată știe că nu va aștepta până în perioada a treia, așa că are de ales în realitate între recompensa din prima perioadă și cea dintr-a doua. În cazul acestei alegeri, recompensa mai timpurie poate avea câștig de cauză.[2] Unii alcoolici susțin că au trecut prin ratio-

---

[1] Putem ilustra acest lucru presupunând că valoarea prezentă a unei unități de utilitate în momentul t în viitor este $1 / (1 + t)$ și că voi suferi din ce în ce mai mult dacă amân vizita mea la dentist: dacă merg azi, sufăr o durere de – 2,75, mâine va fi – 5, iar poimâine va fi – 9. Din perspectiva zilei de azi, valorile prezente sunt – 2,75, $– 5 / (1 + 1) = – 2,5$ și, respectiv, $– 9 / (1 + 2) = – 3$. Prin urmare, poate părea că alegerea optimă din perspectiva zilei de azi este să amân vizita până mâine. Totuși, fiind sofisticat, știu (azi) că mâine valoarea prezentă de a merge mâine la dentist va fi de – 5, iar aceea de a merge poimâine $– 9 / (1 + 1) = – 4,5$, ceea ce determină o preferință de a aștepta până poimâine. Însă, din moment ce azi prefer să merg astăzi mai degrabă decât poimâine, în concluzie mă duc azi.

[2] Putem ilustra cu următorul exemplu: să presupunem că valoarea prezentă a unei unități de utilitate în momentul t în viitor este $1 / (1 + t)$ și că pot beneficia din ce în ce mai mult prin amânarea consumului unei sticle de vin: dacă o beau anul acesta, obțin o plăcere de 2,75, dacă o beau la anul plăcerea va fi 5, iar dacă o beau peste doi ani va fi 9. Din perspectiva zilei de azi, valorile prezente sunt 2,75, $5 / (1 + 1) = 2,5$ și, respectiv, $9 / (1 + 2) = 3$. Prin urmare, poate părea că alegerea optimă din perspectiva primului an este aceea de a bea vinul peste doi ani. Cu toate acestea, pentru că sunt sofisticat, știu (anul acesta) că, la anul, valoarea prezentă a consumării vinului în anul următor va fi de 5, iar aceea de a-l bea peste doi ani va fi $9 / (1 + 1) = 4,5$,

namente asemănătoare: „Știu că voi ceda tentației, prin urmare mai bine cedez imediat." De asemenea, fumătorii naivi care s-au lăsat pentru prima dată pot rezista mai mult decât cei sofisticați, care au încercat de mai multe ori să se lase și care știu care sunt șansele de succes. Cu toate că recăderea în păcatul dependenței nu trebuie neapărat să se datoreze deprecierii hiperbolice, argumentul general este același: dacă poți prezice că vei devia de la cel mai bun plan pe care ți l-ai făcut, atunci vei ajunge să deviezi chiar mai mult sau mai devreme decât atunci când nu ești conștient că vei eșua.

Instrumente extrapsihice

În practică, plănuirea sofisticată împotriva sinelui viitor este, probabil, mai puțin importantă decât gruparea strategiilor și a instrumentelor de preangajare la care mă voi referi în cele ce urmează. Acestea implică afectarea lumii externe, în moduri care nu pot fi anulate imediat și fără costuri, cu scopul de a scădea șansele ca agentul să aleagă recompensa timpurie mai mică în viitor. Cinci strategii ies în evidență: *eliminarea* alegerii recompensei timpurii din setul fezabil, *impunerea unei penalități* pentru alegerea recompensei timpurii, *adăugarea unui bonus* pentru alegerea recompensei târzii, *impunerea unui răgaz* între alegere și livrarea *de facto* a recompensei, precum și *evitarea semnalelor* care ar putea declanșa inversarea preferințelor. Comportamentul de economisire poate ilustra primele patru opțiuni. Dacă încep să economisesc pentru Crăciun, dar văd că scot bani din contul de economii în loc să îi las în pace, atunci pot să mă alătur unui club pentru economii de Crăciun, care nu permite retrageri înainte de data fixată (capitolul 12). Altminteri, aș putea să-mi plasez economiile într-un cont cu o dobândă mare, care prevede o penalitate pentru retragerea mai devreme de data stabilită, combinând astfel penalitatea cu bonificația. Dacă vreau să economisesc pentru bătrânețe, pot fixa un răgaz între decizia de a cheltui economiile și momentul în care fondurile devin disponibile, investind în active nelichide, și nu în acțiuni sau obligațiuni. A cincea opțiune este ilustrată de persoana a cărei poftă pentru desert este declanșată de semnale vizuale. Trucul este de a merge la un restaurant unde desertul nu se servește de pe cărucioare purtate printre mese, astfel încât să fie nevoit să comande din meniu. Putem pune acest lucru în opoziție cu o persoană care are o problemă cu desertul din cauza deprecierii hiperbolice. Pentru această persoană, cea mai bună opțiune este să meargă la un restaurant unde trebuie să comande desertul la începutul mesei.

Oamenii care se înscriu la cursuri de exerciții fizice săptămânale se lasă adesea după o săptămână sau două. Pentru a împiedica acest lucru, ei ar putea (cel puțin teoretic) să semneze un contract cu centrul de fitness, astfel încât să plătească în avans

---

ceea ce determină o preferință de a bea vinul la anul. Din moment ce în primul an, însă, prefer să beau vinul în primul an în loc să îl beau la anul, îl voi bea chiar astăzi.

o taxă dublă de înscriere și să primească înapoi o parte din bani la fiecare ședință la care se prezintă. Oamenii care semnează pentru programe de slăbit pot fi obligați să facă un depozit pe care îl primesc înapoi doar dacă pierd o anumită greutate, stipulată prin contract, uneori cu anexa că, în cazul în care aceștia eșuează, banii urmează a fi donați cauzei politice pe care o detestă. Când am început să predau cursurile care stau la baza cărții de față, mi-am asumat preangajamentul față de studenți că le voi înmâna o schiță a fiecărui capitol la sfârșitul fiecărei săptămâni. Dacă nu mi-aș fi respectat această promisiune, ar fi trebuit să suport costul unei ușoare senzații de ridicol. Dacă îmi e teamă că s-ar putea să-mi anulez ședința la dentist când se apropie momentul să mă duc, îl pot autoriza pe medic să mă taxeze de două ori mai mult decât pe un client obișnuit în cazul în care nu mă prezint. În cazul vinului care se îmbunătățește odată cu trecerea timpului, poți ruga vânzătorul să ți-l păstreze pentru a te proteja de satisfacerea prematură a plăcerii respective. Dacă ți-e teamă că poți citi ultimul roman polițist al autorului tău preferat prea repede, sărind peste paragrafe ca să ajungi la deznodământ, ai putea să cumperi cartea pe casetă audio (și un aparat de redare fără funcție de repede-înainte), care nu îți oferă altă alegere decât aceea de a asculta fiecare cuvânt.

Exemplele din ultimele două paragrafe presupun un preangajament împotriva a două tipuri de tentație: pe de o parte, tentația amânării, care include eșecul de a economisi, de a urma tratamente dureroase sau de a face exerciții fizice sau de a scrie un manuscris, pe de altă parte, tentațiile satisfacției premature, precum aceea de a bea vinul prea repede sau de a sări peste anumite pagini în lectura unei cărți. Aceste tentații își au originea directă în deprecierea hiperbolică. Trecerea timpului este singurul factor implicat. În altă categorie de cazuri, *comportamentul excesiv*, deprecierea hiperbolică poate interacționa cu alte elemente, motivațiile viscerale. Acestea includ: mâncatul în exces, patima jocurilor de noroc și comportamentele de dependență. Poate fi dificil de aflat, în astfel de cazuri, dacă inversarea preferințelor se datorează structurii depreciative sau altor factori. Decizia cuiva de a ține post luată cu burta plină ar putea dispărea pe măsură ce persoana respectivă începe să simtă foamea. Decizia de a renunța la fumat poate să se erodeze la vederea unei alte persoane care își aprinde o țigară. Acesta este fenomenul *dependenței de semnale* – pofte declanșate de semnale vizuale asociate cu consumul substanțelor care provoacă dependență. Decizia de a întrerupe jocurile de noroc determinată de sentimentele de vinovăție resimțite de un jucător față de familie poate fi dejucată odată ce emoția își pierde din tărie (capitolul 8). De asemenea, poate fi dificil să ne dăm seama dacă avem de-a face cu amânare sau cu factori viscerali. O decizie de a lua medicamente regulat poate fi subminată de scăderea tăriei emoției puternice care l-a determinat pe pacient să meargă inițial la doctor.

Odată ce agentul înțelege că este supus acestor alte mecanisme, își poate asuma preangajamentul de a dejuca funcționarea lor. Pentru ca decizia lui de a ține regim să nu fie subminată de foame, acesta poate lua o pastilă care atenuează pofta de mâncare.

Mai drastic, ar putea să își implanteze un dispozitiv care să îi țină maxilarele fixe, astfel încât să nu poată consuma decât hrană lichidă. Dacă știe că pofta sa după desert este dependentă de semnale, acesta poate să nu frecventeze restaurantele unde desertul este servit pe un cărucior care circulă printre mese.[1] Foștii dependenți de heroină se vor feri să meargă în locurile unde obișnuiau să se drogheze. Foștii jucători de ruletă vor învăța să nu meargă la cazinou „doar pentru a-i vedea pe ceilalți cum joacă". Dacă agentul poate prezice că mânia i se va risipi și că, astfel, nu va mai dori să îl pedepsească pe acela care i-a adus ofensa, probabil că va duce la îndeplinire pedeapsa chiar acum. Acest comportament s-a observat în Belgia după război. Pe baza experienței lor de după Primul Război Mondial, mulți belgieni credeau că, după un timp, dorința populară de a impune pedepse severe pentru colaboraționiști va face loc indiferenței. Prin urmare, aceștia voiau ca procesele să se desfășoare cât mai repede cu putință, înainte ca emoțiile să se stingă.

În lupta împotriva dependenței este destul de răspândită strategia de a impune costuri asupra propriei persoane. Când Generalul de Gaulle a vrut să se lase de fumat, le-a spus tuturor prietenilor lui despre intenția sa pentru a crește costurile recidivei. Într-un centru pentru tratarea dependenței de cocaină din Denver, pacienților care sunt medici de profesie li se oferă oportunitatea de a scrie scrisori autoincriminatorii către Consiliul de Stat al Examinatorilor Medicali în care pot mărturisi că utilizează drogul și cer să li se revoce dreptul de a mai profesa. Scrisoarea este trimisă de îndată ce pacientul este depistat pozitiv pentru cocaină. Unii foști alcoolici încearcă să rămână curați utilizând disulfiram, un medicament al cărui efect este acela de a determina stări violente de rău în cazul consumului de alcool.[2]

Amânările autoimpuse pot fi, de asemenea, foarte eficiente pentru a asigura rezistența la pofte. Pentru a mă feri de consumul de alcool din impuls pot să îmi țin băutura într-un seif contorizat. Ca alternativă, aș putea să adopt strategia „nu țin alcool în casă pentru a trebui să merg la magazin să fac rost". Tehnica utilizării de disulfiram combină, de fapt, impunerea de costuri cu amânarea, având în vedere că, odată ce ai luat pastila, trebuie să treacă peste două zile în care să nu o mai iei pentru a te putea apuca din nou de băut fără să ți se facă rău. Centrul pentru tratarea dependenței de cocaină combină, de asemenea, costurile cu amânarea. Acesta le permite oamenilor să iasă din înțelegere printr-o cerere, semnată de un notar, în care se solicită retragerea. Exista o amânare de două săptămâni. Oricine face această cerere pentru retragere

---

[1] Dacă facem o comparație între acest caz și acela în care desertul trebuie comandat la începutul mesei, putem spune că ambele presupun protejarea de efectele *proximității tentației*, fie ea spațială sau temporală.

[2] Forma cea mai obișnuită este administrarea orală. Aceasta funcționează provocând o senzație puternică de rău la consumarea de alcool. Implantarea subcutanată (mai puțin eficientă din punct de vedere fizic, dar eficientă din punct de vedere psihic) este mai puțin întâlnită.

poate recupera scrisoarea autoincriminantă după două săptămâni. Însă dacă, în intervalul celor două săptămâni, retragerea ar fi fost revocată, atunci ar fi fost nevoie de alte două săptămâni pentru o înștiințare. Cu toate că mulți pacienți au invocat procedura de retragere, niciunul nu a rezistat două săptămâni fără să revoce revocarea.

Preocuparea pentru preangajamente în vederea combaterii inconsistenței temporale și a comportamentului excesiv este de dată relativ recentă. Autorii clasici care s-au aplecat asupra subiectului preangajării împotriva *pasiunilor* în sens larg au inclus în aceasta și beția și stările halucinatorii.[1] În *Odiseea*, Homer ne oferă ceea ce a devenit exemplul standard de preangajare: Ulise se leagă de catarg pentru a fi incapabil să reacționeze la cântecul sirenelor. În *Despre mânie*, Seneca scrie: „Cât timp suntem cu mintea limpede, cât timp suntem noi înșine, să cerem ajutorul împotriva unui rău atât de puternic și atât de adesea tolerat de noi. Cei care nu se pot stăpâni la băutură și cărora le e teamă că vor fi imprudenți și insolenți își roagă prietenii să îi îndepărteze din toiul petrecerii; cei care au învățat că nu sunt rezonabili atunci când se îmbolnăvesc ordonă ca, pe timp de boală, să nu li se asculte porunca." În romanul Doamnei de la Fayette *Principesa de Clèves*, prințesa fuge de la Curte pentru a evita tentația de a răspunde avansurilor Ducelui de Nemours; mai târziu, după moartea soțului, când se putea recăsători, aceasta rămâne departe de Curte. „Știind modul în care împrejurările afectează cele mai înțelepte decizii, ea nu era dispusă să suporte riscul de a-și vedea propriile decizii schimbate sau să se întoarcă în locul unde trăise omul pe care l-a iubit." În romanul lui Stendhal *Lucien Leuwen*, doamna de Chasteller se asigură că nu îl va vedea pe Lucien decât în compania unui însoțitor pentru a face prohibitiv de mari costurile pe care le-ar presupune cedarea în fața dragostei pe care acesta i-o poartă.

Aceste strategii sunt destul de des întâlnite. Oamenii pot da foc podurilor pe care le controlează din motive strategice (capitolul 19), însă uneori o fac pentru a nu ceda din cauza fricii. Mă pot feri de petrecerea dată de firmă pentru că știu din experiențele trecute că sunt șanse mari să beau mai mult decât trebuie și că voi avea, din cauza efectului dezinhibant al alcoolului, un comportament agresiv sau amoros pe care mai târziu voi ajunge să îl regret. Sau pot lua decizia de a veni cu partenerul de viață pentru a crește costurile pe care un astfel de comportament le-ar presupune. Simpla *decizie* de a nu bea (sau de a nu deveni prea emoțional dacă o fac) are șanse mai mici să fie eficientă, dată fiind „puterea situației") (capitolul 10). Similar, controlarea mâniei prin instrumentul intrapsihic de a număra până la zece înainte să dai curs exprimării unui răspuns presupune o detașare care tinde să lipsească în vâltoarea pasiunii. Un sfat general pentru a te ajuta pe tine însuți este acela de „a rupe lanțul cât mai devreme", mai degrabă decât să te bazezi pe autocontrol în fața tentației sau provocării. După cum spunea Mark Twain, „este mult mai simplu să nu intri în joc decât să ieși". Pentru un

[1] Fenomenul, menționat în trecere în text, al preangajării în timp ce persoana este cuprinsă de pasiune și se teme că aceasta se va diminua este mult mai puțin întâlnit.

caz extrem mă voi referi la un titlu din *New York Times* (5 aprilie, 1996): „Statul Texas este de acord cu operația în cazul unui violator: urmând să iasă în curând din pușcărie, omul vrea să fie castrat pentru a-și înfrâna nevoia de sex."

Strategiile care se bazează pe amânare par cele mai promițătoare pentru a-i veni de hac iraționalității determinate de emoție. Din moment ce emoțiile tind să aibă un timp de înjumătățire scurt, orice obstacol în executarea imediată a unei tendințe spre acțiune poate fi un remediu eficient. După cum voi observa și mai târziu, autoritățile se bazează pe această trăsătură a emoției când le impun oamenilor un timp de așteptare înainte de a lua anumite decizii importante. Însă sunt rare cazurile acelora care *își* impun amânări cu scopul manifest de a contracara pasiunile. Poate că pur și simplu lipsesc condițiile tehnice necesare. Un exemplu este acela al „căsătoriei-legământ", posibilă în trei state americane (Arkansas, Arizona și Louisiana), o formă opțională de căsătorie în care intri mai greu, dar și ieși mai greu decât dintr-o căsătorie obișnuită. În mod obișnuit, unui cuplu care a intrat într-o astfel de căsătorie-legământ i se poate acorda divorțul doar după doi ani de despărțire, spre deosebire de timpul normal de așteptare, de șase luni. Mica minoritate (mai puțin de un procent din cuplurile care se căsătoresc) care folosesc această opțiune o fac de bună seamă pentru a semnala angajamentul pe care îl iau unul față de celălalt și pentru a se proteja de tentațiile și pasiunile efemere.

Preangajarea implică de multe ori ajutorul celorlalți, organizații sau autorități publice. Acestea trebuie, însă, să fie independente de agentul care dă instrucțiunile de preangajare din moment ce, altminteri, acesta ar putea la fel de bine să le revoce. Pentru a lupta cu dependența de opiu, Samuel Coleridge a angajat pe cineva care să folosească forța dacă îl prinde că încearcă să intre în vreo farmacie. Când omul a încercat să i se opună, Coleridge i-a zis: „Ah, prostii! E o urgență, o urgență îngrozitoare la care chiar nu mă așteptam. Nu are importanță ce ți-am zis în trecut. Ce îți spun *acum* este că, dacă nu mă lași să intru în această farmacie respectabilă, voi avea un temei foarte bun ca să te dau în judecată pentru agresiune." Asemănător, Mao Zedong a dat ordin ca orice comenzi pe care le-ar putea da după ce ia somnifere să fie ignorate. Când, după ce a luat pastilele, i-a ordonat unui consilier să trimită o invitație echipei de tenis de masă americane să viziteze China (începutul relațiilor chino-americane) și consilierul l-a întrebat: „Să țin cont de cuvintele dumneavoastră?", acesta i-a răspuns: „Da, ține cont. Fă-o repede. Dacă nu, n-o să mai fie timp."

Organizațiile sunt instrumente mult mai de încredere pentru autocenzură. Clinica de dezintoxicare din Denver și cluburile de Crăciun oferă opțiuni de autocenzură pe care indivizii nu le-ar fi putut inventa singuri și care au fost proiectate deliberat pentru a acorda ajutor în combaterea anumitor probleme[1] *și* pentru a împiedica revocarea

[1] Seifurile cu temporizare, dimpotrivă, nu au fost făcute pentru a-i face pe oameni să se împotrivească problemelor cu băutura.

instrucțiunilor date. În Norvegia, Legea protecției sănătății psihice le permite indivizilor să se angajeze *voluntar, dar ireversibil* într-un tratament de trei săptămâni într-o instituție psihiatrică. Însă se pare că sistemul nu funcționează pentru că doctorii au dreptul, dar nu și obligația, să-i rețină pe indivizii intrați în clinică. Pentru a impune eficiență, pacienților ar trebui să li se permită să dea în judecată spitalul în cazul în care, la cererea lor, sunt externați prematur.

În 1996, statul Missouri a demarat un program de autoexcludere pentru împătimiții jocurilor de noroc. Oricine semnează pentru a fi pus pe lista de autoexcludere primește o interdicție pe viață de a intra în orice ambarcațiune care găzduiește un cazinou din Missouri. Dacă persoana încearcă să ignore interdicția și să joace în vreun cazinou, aceasta trebuie îndepărtată, iar „acela care deține licența de a administra un cazinou va coopera cu comisionarul în vederea raportării incidentului autorităților competente, care vor pune persoana sub acuzare (...) pentru încălcarea proprietății, care este o faptă penală". Jucătorii care s-au autoexclus nu se pot bucura de câștigurile lor dacă reușesc cumva să intre într-un cazinou, joacă și câștigă. Statul poate, de asemenea, să își asume un rol mai activ, impunând amânări ale avortului, ale achiziționării armelor de foc sau ale divorțului (și ale căsătoriei!) și acordându-le consumatorilor o perioadă de trei zile sau de o săptămână pentru ca aceștia să poată anula anumite cumpărături făcute într-un moment de entuziasm. Însă acestea sunt cazuri de *paternalism al statului*, și nu de *autopaternalism asistat de stat.*

Uneori, constituțiile politice sunt considerate instrumente de preangajare sau o formă de autopaternalism colectiv. John Potter Stockton scria în 1871: „Constituțiile sunt lanțuri de care cetățenii se leagă în momente de limpezime a minții pentru a nu muri de o mână sinucigașă în momentele de nebunie." Constituțiile pot fi caracterizate și prin altă metaforă, și anume că sunt legături impuse de Petru atunci când e treaz pentru Petru atunci când e beat. Bicameralismul este citat adesea drept exemplu de preangajament politic: impunând ca toată legislația să treacă prin două Camere se dă un răgaz, astfel încât pasiunile impulsive să se domolească, iar rațiunea (sau interesul!) să revină în avantaj. Impunerea unor amânări asupra amendamentelor constituționale a fost justificată prin același argument. Cu toate acestea, dacă înțelegem preangajarea drept *auto*-cenzură, atunci extinderea de la cazul individual la cel general – altfel spus, de la cazul intragenerațional la cel intergenerațional – este destul de dubioasă. În loc să avem de-a face cu comunități care se cenzurează, avem de-a face cu majorități care cenzurează minorități și generații prezente care le cenzurează pe cele viitoare. Mai mult, din moment ce constituțiile sunt scrise în vremuri tulburi, fondatorii lor sunt adesea cuprinși de pasiune. Fiind „beți", acestora le poate scăpa din vedere să ia precauții împotriva beției. Astfel, în 7 septembrie 1789, când Adunarea Constituantă purta o dezbatere despre unicameralism *versus* bicameralism cu privire la noua Constituție, deputatul Adrien Duquesnoy scria următoarele în jurnalul său: „Dacă ar fi să facem o evaluare de probabilitate, pare foarte clar că majoritatea Adunării nu va vota niciodată

pentru bicameralism. Acest rezultat poate avea foarte multe dezavantaje, dar aceasta este starea de fapt, iar mințile sunt atât de exaltate încât niciun alt rezultat nu este posibil; poate că va fi posibilă o schimbare în câțiva ani. Vom ajunge să înțelegem că o adunare unică pentru o națiune atât de impetuoasă precum a noastră poate produce efectele cele mai groaznice."

◈◈◈

Notă bibliografică

Instrumentul intrapsihic al grupării opțiunilor a fost discutat pe larg de G. Ainslie, în special în *Picoeconomics* (Cambridge University Press, 1992). O încercare de a furniza fundamente raționale „regulilor private" se regăsește în R. Bénabou și J. Tirole, „Willpower and personal rules", *Journal of Political Economy*, 112 (2004), pp. 848-886. Discuția privitoare la miopia în decizie se bazează pe O.-J. Skog, „Hyperbolic discounting, willpower, and addiction", în J. Elster (ed.), *Addiction: Entries and Exits* (New York, Russell Sage Foundation, 1999). Răspunsuri strategice ale unor indivizi sofisticați care nu sunt conștienți de înclinația lor de a deprecia viitorul hiperbolic sunt discutate de T. O'Donoghue și M. Rabin, „Doing it now or later", *American Economic Review*, 89 (1999), pp. 103-124. Ideile de preangajare și autocenzură pentru a face față înclinațiilor iraționale sunt discutate de T. Schelling în „Egonomics, or the art of self-management", *American Economic Review: Papers and Proceedings*, 68 (1978), pp. 290-294, și în câteva dintre publicațiile sale ulterioare. Am discutat aceste aspecte în *Ulysses and the Sirens*, ediție revizuită (Cambridge University Press, 2000), precum și în „Don't burn your bridges before you come to it: Ambiguities and complexities of precommitment", *Texas Law Review*, 81 (2003), pp. 1751-1788. O abordare exhaustivă a eșecului de a lua medicamente date pe rețetă este cartea lui G. Reach, *Pourqui se soigne-t-on* (Paris, Editions de Bord de l'Eau, 2005). Povestea despre Coleridge se găsește în Thomas de Quincey, *Confessions of an Opium Eater* (Londra, Penguin Books, 1968), p. 145. Povestea despre Mao Zedong se găsește în J. Chang și J. Halliday, *Mao: The Unknown Story* (New York, Knopf, 2005), pp. 580-581. Dacă în *Ulysses and the Sirens* eram entuziasmat de ideea conform căreia constituțiile pot fi instrumente de preangajare, în *Ulysses Unbound* m-am răzgândit.

Capitolul 14

CÂTEVA IMPLICAȚII ASUPRA INTERPRETĂRII TEXTELOR

Există un punct de vedere larg răspândit conform căruia activitatea științifică are trei părți distincte sau ramuri: științele umane, științele sociale și științele naturii. În anumite cazuri, acest mod de a compartimenta știința poate fi util, însă, în altele, aceste distincții rigide pot sta în calea interdisciplinarității. În Partea a IV-a voi susține că științele sociale pot beneficia de descoperirile din studiul biologic al omului și al altor animale. În acest capitol, voi argumenta că științele umane și științele sociale au mai multe în comun decât se presupune de obicei. În special, voi încerca să arăt că *interpretarea* operelor de artă și *explicația* sunt activități foarte apropiate. A înțelege o operă de artă înseamnă a o explica în termenii stărilor mentale anterioare ale creatorului. O operă de artă de *succes* este una căreia i se poate da o explicație bazată pe alegerea rațională. În același timp, voi argumenta împotriva a ceea ce am putea numi „interpretarea prin consecințe", expresie pe care urmează să o clarificăm. Abordarea pe care o propun în cele ce urmează nu acoperă toate formele artistice. Chiar și în literatură, artă la care mă voi limita, are sens doar pentru unele romane și piese de teatru clasice (scrise înainte de 1850), definite prin convenția tacită că evenimentele și personajele descrise *ar fi putut fi adevărate.*[1]

Să luăm mai întâi în considerare raționalitatea ca motiv al *personajelor* din romane sau piese de teatru. O problemă clasică în critica literară și pentru care s-au dat multe explicații este de ce amână Hamlet să răzbune moartea tatălui său. Unele explicații fac apel la iraționalitate, la slăbiciunea voinței și la depresia clinică. Există, însă, o abordarea simplă care face apel la alegerea rațională. Cu toate că, la început, Hamlet crede lucrurile pe care i le-a spus fantoma tatălui său despre Claudius, acesta se decide ulterior să *culeagă mai multe informații* prin înscenarea unei piese de teatru cu scopul de a „da în vileag conștiința regelui". Odată ce i se confirmă credința prin reacțiile

[1] Un exemplu timpuriu de încălcare a acestei convenții se produce spre sfârșitul piesei lui Ibsen *Peer Gynt* (1867), când lui Peer îi este teamă de înec, iar „trecătorul misterios" îi spune că „nu se moare la mijlocul actului V".

regelui, lui Hamlet îi lipsește *oportunitatea* de a-și îndeplini dorința, anume aceea de a-l face pe Claudius să ardă pe vecie în focurile iadului. Cu toate că are o ocazie să îl ucidă pe Claudius în timp ce acesta se roagă, comiterea unui astfel de act i-ar aduce lui Claudius, conform învățăturilor teologice contemporane, mântuirea mai degrabă decât damnarea. Mai târziu, îl ucide pe Polonius, aflat în spatele unei perdele, crezând *în mod greșit, dar nu irațional*, că acesta este regele. Date fiind informațiile de care dispunea, credința lui că regele se ascunde în spatele perdelei era rațională. Mai mult, Hamlet nu avea vreun motiv să adune *mai multe* informații întrucât putea presupune în mod rezonabil că persoana care se ascundea în spatele perdelei în prezența reginei trebuie să fi fost regele.

Nu vreau susțin că aceasta este interpretarea corectă (de fapt, nu am spus încă nimic cu privire la ce înseamnă că o interpretare este „corectă"). Argumentul pe care vreau să îl dau este simplu: cele trei episoade pe care le-am menționat sunt, la o privire superficială, coerente cu ideea că Hamlet își urmărește în mod rațional scopul răzbunării morții tatălui său. O altă întrebare este aceea dacă ideea este coerentă cu acuzațiile repetate pe care Hamlet și le aduce sieși, conform cărora îi lipsește capacitatea luării unei decizii ferme de a se răzbuna. Mulți comentatori interpretează aceste monologuri faimoase ca semne ale slăbiciunii voinței și consideră că primele două episoade sunt bazate pe scuze pentru inacțiune, pornite din autoamăgire. (Cel de-al treilea episod este mai greu de împăcat cu această viziune.) Acum, cu toate că slăbiciunea voinței și autoamăgirea violează canoanele raționalității, ele sunt perfect *inteligibile* (capitolul 3). Când avem de-a face cu dezvoltarea internă a unei opere de artă, inteligibilitatea, și nu raționalitatea, constituie ideea cea mai utilă pentru scopul interpretării.

Ca alternativă la punctul de vedere intern, ne putem orienta către punctul de vedere extern, al autorului. La întrebarea „De ce amână Hamlet să se răzbune până în actul cinci?" putem răspunde: „Moartea regelui trebuie să aibă loc la finalul piesei."[1] Avem de-a face cu o chestiune care ține de construcția dramatică, și nu de psihologie. În sine, un astfel de răspuns nu are cum să fie satisfăcător. Dacă Shakespeare ar fi tărăgănat tema răzbunării printr-o serie de evenimente arbitrare sau coincidențe ad-hoc doar în scopul amânării evenimentului pentru finalul piesei, am fi considerat piesa un eșec. Mai mult, ar fi fost un caz de *iraționalitate auctorială*.

Raționalitatea auctorială este asemănătoare raționalității atribuite lui Dumnezeu. La fel ca Dumnezeu, autorul pune în mișcare un proces prin care fiecare eveniment

[1] Spre deosebire de cuvintele pe care le pune Ibsen în gura „trecătorului misterios" (vezi nota precedentă), Shakespeare nu îl poate face pe Hamlet să spună: „Nu îl pot ucide pe rege înainte de actul V".

poate fi *explicat în două moduri*, cauzal și teleologic. Preiau această idee de la Leibniz, care scrie că există

> două imperii, unul al cauzelor eficiente, celălalt al cauzelor finale, fiecare dintre ele fiind suficient în detaliu pentru a da o rațiune pentru întreg, ca și cum celălalt nici nu ar exista. Dar niciunul nu este adecvat fără celălalt dacă luăm în considerare originea lor, întrucât acestea emană din unica sursă unde se găsesc împreună puterea care face cauzele eficiente și înțelepciunea care stabilește cauzele finale.

Scopul lui Dumnezeu este acela de a crea cea mai bună dintre lumile posibile. Enunțată astfel încât să includă și dimensiunea temporală, ideea poate fi înțeleasă drept *cea mai bună dintre secvențele posibile*. Cu toate că trecerea de la o stare a universului la următoarea are loc prin cauzalitate fizică obișnuită, starea inițială a universului și legile cauzalității au fost alese astfel încât să maximizeze perfecțiunea globală a secvenței.

Dacă ne limităm la drama clasică sau la romanul clasic, sarcina autorului este aceea de a dezvolta scenariul prin ceea ce personajele spun și fac, de cele mai multe ori ca răspuns la ceea ce spun și fac celelalte personaje. Scopul este de a prezenta totul într-un mod care maximizează valoarea estetică. Astfel, fiecare acțiune sau enunț al unui personaj poate fi explicat în două moduri, atât ca reacție la acțiuni și enunțuri precedente (sau evenimente externe), cât și ca un generator de surpriză, tensiune și, în cele din urmă, rezolvare a tensiunii la nivelul cititorului. Prima explicație se bazează pe inteligibilitatea personajelor, a doua pe raționalitatea autorului într-un sens pe care voi încerca să îl clarific în cele ce urmează.

Faptul că autorii alcătuiesc, de cele mai multe ori, mai multe schițe sau versiuni preliminare ale operei lor înainte să fie mulțumiți sau înainte să pună punct este o dovadă incontestabilă că aceștia sunt angajați într-un proces de *alegere* și că dețin criterii implicite sau explicite pentru *ceea ce este mai bun*. Faptul că aceste versiuni preliminare implică, de obicei, variații *minore* sugerează că autorii țintesc un punct de *maximum local* al oricărei forme de mai bine pe care năzuiesc să o atingă. Totuși, diferența dintre un autor și cineva care doar urcă o pantă este aceea că, în cazul celui dintâi, creativitatea trece dincolo de simpla alegere.[1] Motivul pentru care creația unei opere literare nu poate fi redusă la alegerea rațională este că numărul de secvențe de cuvinte cu sens este mult prea mare pentru a fi scanat de o singură persoană care să selecteze secvența considerată ca fiind „cea mai bună". Cu toate că un „creator rațional" poate încerca să facă problema mai ușor de rezolvat excluzând în mod deliberat anumite secvențe (aceasta este una dintre funcțiile măsurii și rimei în poezie), vor rămâne de obicei

[1] Din același motiv, încercările de a găsi explicații ale schimbărilor tehnice prin alegerea rațională sunt sortite eșecului.

prea multe opțiuni pentru ca alegerea să fie un mecanism fezabil de selecție. În locul acestui mecanism, autorul va trebui să se bazeze pe mașinăria asociativă inconștientă.

Prin urmare, creația rațională presupune în mare parte, a găsi a doua zecimală corectă sau, pentru a folosi altă metaforă, a urca pe vârful celui mai apropiat deal. Tot printr-o metaforă, putem spune că avem de-a face cu ceva care ține de emisfera stângă. Sarcina atribuită emisferei stângi, de a nimeri prima zecimală corectă sau de a găsi un deal care este mai înalt decât celelalte, nu ține de raționalitate. Însă, chiar redusă doar la sarcina unor reglaje fine, raționalitatea auctorială contează. După cum sugerează expresia „o capodoperă minoră", mai bine să ajungi în vârful unei coline decât să rămâi împotmolit pe panta unui munte. Fără să vreau să aduc în discuție chestiuni care țin de valoarea relativă, *Cronica unei morți anunțate* și *Privește, înger, către casă* pot ilustra cele două posibilități.

În cele ce urmează, voi enumera și voi supune dezbaterii câteva cerințe pe care raționalitatea le impune autorului. În primul rând, acțiunile și enunțurile personajelor trebuie să fie inteligibile. În al doilea rând, autorul trebuie să satisfacă cerințele îngemănate ale *deplinătății* și *parcimoniei*. În al treilea rând, opera trebuie să curgă *în aval*, în sensul că trebuie redusă la minim utilizarea de accidente și coincidențe. În al patrulea rând, opera trebuie să ofere o înșiruire satisfăcătoare din punct de vedere psihologic de acumulări și rezolvări de tensiune.

Inteligibilitatea poate fi absolută sau relativă, iar, dacă este relativă, poate fi globală sau locală. Chestiunea deplinei inteligibilități echivalează cu posibilitatea ca orice ființă umană să se poată comporta în modul descris. Chestiunea inteligibilității relative globale vizează coerența comportamentului unei persoane fictive cu caracterul acesteia, așa cum apare descris în general în cadrul operei. Chestiunea inteligibilității relative locale vizează coerența comportamentului unei persoane fictive cu comportamentul său în situații similare întâlnite în operă. Pe când cerințele de inteligibilitate locală absolută și relativă sunt constrângeri cruciale pentru raționalitatea auctorială, inteligibilitatea relativă globală nu este astfel. Ba chiar respectarea unei astfel de cerințe ar putea fi văzută ca un defect din punct de vedere estetic.

În anumite cazuri, inteligibilitatea absolută poate fi încălcată de un exces de raționalitate. Ne putem referi din nou la *Medeea* lui Euripide sau la *Fedra* lui Racine, ambele personaje în egală măsură lucide cu privire la patimile lor autodistructive. Acestea sunt portretizate ca suferind de o slăbiciune a voinței în sens strict, știind că ceea ce fac contravine judecății pe care acestea o au *chiar în momentul acțiunii*. Cu toate că pasiunea le determină să devieze de la acea judecată, judecata în sine nu este afectată de pasiune. Hermiona a lui Racine este un personaj mult mai credibil. Pentru că judecata ei este încețoșată de emoție, aceasta se autoamăgește, dar nu manifestă o voință slabă. Sugestia mea – nu este nimic mai mult – este că prezența simultană a unor emoții

extreme și a unei lucidități cognitive nu corespunde deloc cu ceea ce cunoaștem despre natura umană.

Dacă prea multă rațională poate fi ininteligibilă, irațională poate fi perfect inteligibilă. Ce poate fi mai inteligibil decât reacția domnului de Rênal, din *Roșu și negru* de Stendhal, care, în fața semnelor clare că soția sa are o relație cu Julien Sorel, alege să creadă în fidelitatea ei? Dorința este mama gândului. Mai paradoxale sunt acele cazuri în care dorința ca soția să fie fidelă determină credința că nu este, împotriva dovezilor. În *Othello*, „pentr-un gelos, și-un fulg ușor ca vântul e o dovadă tot atât de tare ca slova din scripturi". Primul este un caz de scurtcircuitare, al doilea – unul în care firele se încrucișează (capitolul 3).

Inteligibilitatea relativă, încălcată de o persoană dintr-o piesă de teatru sau dintr-un roman care joacă „nepotrivit cu caracterul său", ridică o serie de probleme. În primul rând, trebuie să ținem cont de argumentele psihologilor, conform cărora trăsăturile de caracter tind să fie locale, și nu generale (capitolul 10). În timp ce mulți autori (Hamsun îl menționează pe Zola) subscriu la psihologia populară, care favorizează coerența între situații, autorii buni (aici este menționat Dostoievski) nu respectă această exigență, dezamăgindu-și cititorii, care se așteaptă ca personajul să se comporte „în conformitate cu caracterul său"; însă nu aceasta este publicul vizat de opera literară. După cum vom vedea în cele ce urmează, chiar și autorii buni pot fi constrânși de psihologia defectuoasă a cititorilor lor, dar credința în trăsături globale nu este una pe care să trebuiască să o respecte. Cititorii au, însă, și ei dreptul de a se aștepta la coerență locală. Dacă se vâră singur în bucluc, astfel încât, pentru dezvoltarea intrigii, este obligat să îngăduie unui personaj să se comporte într-o manieră incoerentă local, autorul violează contractul implicit cu cititorul. Acțiunea trebuie să se dezvolte așa cum curge apa unui râu, pe firul ei natural, în aval; ea nu poate fi forțată de autor să curgă în amonte.

Voi ilustra această idee recurgând la unele dintre comentariile marginale ale lui Stendhal făcute în manuscrisul ultimului său roman, terminat și publicat postum sub titlul *Lucien Leuwen*. Personajul lui Stendhal se îndrăgostește de o tânără văduvă, doamna de Chasteller. Sentimentele acestuia sunt împărtășite, dar acesta nu îndrăznește să îi facă avansuri. Însăși delicatețea firii sale, care îl face superior „celui mai desăvârșit Don Juan" și, prin urmare, capabil să inspire dragoste, îl face inferior oricărui „tânăr parizian prost-crescut", care ar ști imediat cum să facă față situației. Pentru a duce intriga mai departe, Stendhal trebuie să își aducă personajele laolaltă, dar nu prea știe cum să facă asta. Pe marginea manuscrisului stă scris: „Cum spune cronicarul: nu te poți aștepta de la o femeie virtuoasă să se arunce în brațele cuiva; ea trebuie convinsă. Cel mai bun ogar nu poate face mai mult decât să aducă vânatul în bătaia puștii. Dacă vânătorul nu trage, câinele este neajutorat. Romancierul seamănă

cu câinele." Comentariul ilustrează frapant nevoia ca personajele unui roman să se comporte „conform caracterului lor".

În cele din urmă, Stendhal reușește să găsească o situație în care dragostea lui Lucien pentru doamna de Chasteller și a acesteia pentru el să poată fi arătate și înțelese, dar nu încă declarate. Însă dificultățile sale nu se opresc aici. Planul lui Stendhal pentru roman urma rețeta hollywoodiană de tip dialectic: băiatul întâlnește o fată, băiatul și fata se despart, băiatul și fata ajung în cele din urmă împreună. După cum am văzut, Stendhal avea probleme cu stabilirea tezei. Pentru a produce antiteza, acesta se folosește de un instrument ridicol și, evident, teleologic, făcându-l pe Lucien că creadă că doamna de Chasteller, pe care o vede zilnic de aproape, a dat naștere pe neașteptate unui copil. Însă obstacolul cel mai mare avea să fie sinteza. Cu toate că nu știm de ce nu a apucat să scrie partea a treia, în care îndrăgostiții ajung iar împreună, putem specula că uniunea lor nu ar fi plauzibilă. În partea a doua a romanului, după despărțire, Lucien devine destul de cinic, onest din punctul de vedere al standardelor laxe ale Monarhiei din Iulie, însă foarte diferit de persoana ciudat de delicată de care se îndrăgostise doamna de Chasteller. Probabil că Stendhal s-a gândit că faptul de a o face pe aceasta să îl iubească pe Lucien după transformare ar încălca inteligibilitatea relativă.

Aristotel scrie că „povestea (...) trebuie să reprezinte o singură acțiune, un întreg complet în care incidentele separate să fie atât de puternic legate între ele, încât transpunerea sau eliminarea unui element ar dezmembra și disloca întregul. Acel ceva care nu aduce nicio diferență perceptibilă prin prezența sau absența sa la întreg nu este o adevărată parte a întregului." Putem interpreta acest pasaj ca fiind expresie a idealurilor deplinătății și parcimoniei. Cititorul are dreptul să creadă că autorul i-a prezentat toate informațiile necesare pentru o bună înțelegere a dezvoltării acțiunii.[1] Invers, cititorul are dreptul să se aștepte ca autorul să îi spună că plouă când un personaj pleacă de acasă doar dacă acest element are o relevanță ulterioară și are dreptul să creadă că un discurs atribuit unui personaj este prezent cu intenția de a spune ceva despre persoana respectivă sau de a servi ca premisă pentru acțiunea altor personaje.[2]

---

[1] Firește, detalii potențial relevante pot fi omise pentru a face loc imaginației cititorului. Creația rațională este compatibilă cu (și poate chiar impune) ca unele spații libere să fie completate de cititor. Însă, dacă artistul supraestimează imaginația publicului său, efortul poate fi considerat un eșec. Să presupunem că un romancier încearcă să sugereze incompatibilitatea de temperament a eroinei cu eroul făcând numerele de pe strada unde locuiesc prime între ele, astfel încât să nu aibă divizori comuni. Cu excepția unor circumstanțe excepționale, autorul nu se poate baza pe faptul că cititorul va fi în măsură să își dea seama de un astfel de detaliu.

[2] Sigur, redundanța nu este întotdeauna un lucru de evitat din moment ce poate avea o funcție estetică. Pentru a reda plictisul, redundanța poate fi mult mai eficientă decât enunțul făcut de autor. Însă chiar și atunci nu ar avea sens dacă o repetiție ar plictisi *cititorul* și nu ar evoca plictisul resimțit de un personaj.

Mai devreme m-am referit la caracterul „descendent" al unei intrigi reușite folosind drept exemplu comportamentul unui personaj în „conformitate cu caracterul". În general, intrigile reușite nu ar trebui să se bazeze pe evenimente puțin probabile, pe accidente și coincidențe. În *Middlemarch*, întâlnirea dintre Raffles și domnul Bulstrode – un element crucial în dezvoltarea poveștii – are un aspect atât de forțat încât cursul, altminteri fluid, al romanului începe să deraieze. Sigur, și accidentele își au locul lor într-un roman. Moartea accidentală a unui părinte poate declanșa sau modela dezvoltarea acțiunii, la fel cum o poate face moartea ambilor părinți în același accident. Dar, dacă dezvoltarea are nevoie de moartea lor în *două* accidente diferite, atunci credibilitatea este pusă la încercare. Moartea convenabilă a partenerului de viață, care îngăduie eroului sau eroinei să se căsătorească cu adevărata sa dragoste, este, de asemenea, un semn al inerției de condamnat a autorului.

Însă psihologia cititorilor nu rezonează atât de puternic cu teoria probabilităților. Să presupunem că autorul are de ales, pentru a ajunge din punctul A al dezvoltării în punctul B, între doi și, respectiv, șase pași. Ca să fie mai clar exemplul, să presupunem că scenariul în doi pași presupune evenimente a căror șansă să se întâmple este de 0,9 și 0,2, în timp ce scenariul din șase pași presupune că toate evenimentele se întâmplă cu o probabilitate de 0,75. Presupunând că evenimentele din fiecare șir sunt independente unele de altele, șirul din doi pași este mult mai probabil să se întâmple (0,18 *versus* 0,178), însă șirul din șase pași va fi considerat ca având proprietatea de a curge în aval. Plauzibilitatea globală a unui scenariu depinde mai mult de plauzibilitatea verigii celei mai slabe decât de numărul de legături. Cred că un autor ar trebui să respecte acest capriciu particular al cititorilor săi din moment ce, în acest mod, nu trebuie să facă apel la coincidențe facile, însă puțin probabile.

Chiar și un flux descendent poate avea multe meandre înainte să ajungă să se verse calm undeva, în mare. Prin urmare, autorul este obligat să furnizeze surprizele necesare pentru cititori și spectatori, precum și obstacolele pentru personaje, astfel încât publicul să își mențină viu interesul pentru poveste. Repertoriul acestor stratageme este uriaș, prea uriaș pentru a fi trecut în revistă sau clasificat. Unele dintre strategii sunt strâns legate de gen. În teatru, comedia, drama și tragedia au mijloace diferite la dispoziție. În timp ce comedia se poate baza adesea pe *neînțelegeri* pentru a genera tensiuni, drama și, chiar într-o măsură mai mare, tragedia se pot baza pe *ignoranță*. Pe măsură ce neînțelegerile se risipesc, ele fac loc bucuriei; pe măsură ce ignoranța se diminuează, apare dezastrul. Romancierii își pot adăuga propria voce la aceea a personajelor pentru a genera incertitudine, atâta timp cât nu induc în eroare cititorii în mod deliberat.

Sunt acum în măsură să spun ce înțeleg prin „interpretarea corectă" a unui text. După cum am menționat la început, este o chestiune care ține de explicație. Din moment ce toate explicațiile sunt cauzale (inclusiv acelea care se referă la intenții ca fiind

cauze) și din moment ce o cauză trebuie să preceadă efectul, rezultă că percepțiile *actuale* ale publicului în raport cu o operă sunt strict irelevante. Percepțiile intenționate, dimpotrivă, pot fi parte a explicației. Printre cauzele antecedente ale unei opere de artă, intenția auctorială nu este tot ce contează. Atitudinea inconștientă a autorului poate avea, de asemenea, o influență. Astfel, *Insula misterioasă* a lui Jules Verne e posibil să fi fost modelată atât de intențiile sale antirasiste, cât și de prejudecățile sale rasiste. Mă voi referi în cele ce urmează strict la intențiile conștiente pentru a simplifica problema.

Interpretarea unei opere literare presupune, prin urmare, că trăsături importante ale operei pot fi puse pe seama unor decizii pe care le-a luat autorul cu scopul de a spori valoarea estetică a experienței pe care un public anume este de așteptat că o va căpăta în urma parcurgerii operei respective. Pentru a decripta aceste mecanisme, criticii literari trebuie să urmeze aceleași proceduri pe care le urmează alți oameni de știință. Aceștia pot face apel la versiuni preliminare ale operei, dacă acestea există, și pot să se bazeze pe declarații ale autorului însuși cu privire la operă, ca, de exemplu, adnotările lui Stendhal. Criticii pot face apel la alte opere ale aceluiași autor pentru a vedea dacă se pot desprinde anumite caracteristici asemănătoare cu privire la alegerile făcute. Ei se pot referi la opere contemporane pentru a distinge convențiile care încadrează alegerile de alegerile în sine și pot, de asemenea, să utilizeze alte surse contemporane pentru a determina așteptările publicului care, probabil, l-au constrâns pe autor.

Procedând în acest mod, metoda lor nu este cu nimic diferită de cea a altor istorici. Asemenea istoricilor, criticii se confruntă cu problema finitudinii datelor, având în vedere că trecutul nu poate fi supus unor experimente. La fel ca istoricii, aceștia pot încerca să reducă la minimum tentația de a „extrage cunoștințe din date" prin extrapolarea unor surse vechi, căutând surse noi și derivând implicații noi din interpretările care urmează a fi testate în raport cu probele disponibile. Spre deosebire de alți istorici, interpretarea lor este foarte adesea, deși nu întotdeauna, însoțită de *judecăți de valoare*. Oare autorul a reușit în ce și-a propus ori s-a apropiat mai degrabă de succes decât de eșec în scopul său de a crea un maximum local de valoare estetică? Unii scriitori, firește, nu au acest scop. Ei pot fi mai degrabă interesați de câștig sau de scrierea operelor de propagandă, scopuri care au condiții diferite de raționalitate. Însă, dacă putem argumenta în mod plauzibil ipoteza că autorul a urmărit pretenții estetice, atunci e logic să ne întrebăm, la fel cum o facem în raport cu orice alt scop, cât de mult au fost atinse aceste scopuri.

Am spus mai devreme că eșecurile auctoriale pot fi inteligibile. Am argumentat că autorii sunt supuși unei duble presiuni: trebuie să dezvolte acțiunea, ceea ce se poate face doar prin acțiuni inteligibile și enunțuri ale personajelor. Putem să-i învinovățim dacă sacrifică cel din urmă scop în favoarea primului – adică, dacă renunță

la cauzalitate în favoarea teleologiei –, însă tot putem *înțelege* de ce au făcut așa ceva. Chiar dacă este implauzibilă din punct de vedere cauzal, putem face inteligibilă din punct de vedere teleologic amânarea lui Hamlet, în virtutea faptului că Shakespeare avea nevoie de această amânare a răzbunării până la finalul piesei. Și aceasta este tot o interpretare. Cu toate că, evident, este foarte diferită de o interpretare a amânării care recurge la psihologia lui Hamlet și la circumstanțe, ea oferă răspuns la aceeași întrebare: care este motivul amânării? Deși într-o operă literară de bună calitate totul se poate explica în două moduri, operele imperfecte pot admite doar o singură interpretare.

Voi încheia citând un exemplu despre cum interpretarea poate să încalce sau să ignore condițiile explicației. Câțiva autori contemporani au susținut că Fanny Price din *Mansfield Park* este un personaj uneltitor, care își urmărește doar propriile scopuri, modestia sa fiind doar o stratagemă pentru a-l câștiga pe Edmund Bertram. Mai mult, aceștia au argumentat chiar că numele personajului sugerează „sexul pentru bani". Aceste afirmații *eșuează dacă sunt supuse la două teste de intenționalitate.* În primul rând, nu există vreo dovadă în roman pentru a-i atribui intenții uneltitoare lui Fanny Price. Cu toate că modestia sa este, de fapt, răsplătită, acea *consecință* a comportamentului ei nu poate să o explice.[1] În al doilea rând, nu există dovezi care să-i atribuie lui Jane Austen intenția de a-i face pe cititori să o vadă pe Fanny Price ca un fel de semi-prostituată. Cu toate că textul poate provoca anumite asocieri în mintea unor cititori, criticii pe care i-am menționat mai sus nu oferă vreo dovadă conform căreia Austen a intenționat să producă în mintea cititorilor asocierea numelui „Fanny" cu eroina din romanul pornografic *Fanny Hill* sau a numelui „Price" cu plata pentru sex. Aceste „interpretări prin consecințe" au multe în comun cu explicațiile funcționale din științele sociale, care se bazează pe metode arbitrare constrânse nu de fapte, ci doar de limitele ingenuității oamenilor de știință care le propun. În Partea a IV-a, însă, ne vom ocupa de unele tipuri respectabile ale „explicației prin consecințe".

◈◈◈

Notă bibliografică

Abordarea generală pe care o folosesc în acest capitol este adesea acuzată de o „eroare intenționalistă". Sunt de acord cu răspunsul dat de N. Carroll și cu critica sa, în special în „Art, intention and convention", în G. Iseminger (ed.), *Intention and Interpretation* (Philadelphia, Temple University Press, 1992), și în „The intentional fallacy: Defending myself," *Journal of Aesthetics and Art Criticism*, 55 (1997), pp. 305-309. În

[1] De asemenea, ipoteza unei Fanny Price cu comportament de mercenar nu poate da seamă de motivele pentru care a respins propunerea de căsătorie a înstăritului Henry Crawford.

„Hermeneutics and the hypothetico-deductive method", în M. Martin și L. McIntyre (editori), *Readings in the Philosophy of the Social Sciences* (Cambridge, MA, MIT Press, 1994), D. Føllesdal oferă o interpretare la *Peer Gynt* asemănătoare cu cea propusă de mine, doar că această piesă nu este constrânsă de convenția conform căreia evenimentele și personajele descrise ar fi putut fi reale. Observația că amânarea lui Hamlet poate fi rodul unor preocupări dramaturgice i-o datorez lui E. Wagenknecht, „The perfect revenge – Hamlet's delay: A reconsideration", *College English*, 10 (1949), pp. 188-195. Am discutat ideea că operele de artă sunt puncte de maximum local în capitolul 3 din *Ulysses Unbound* (Cambridge University Press, 2000). Acel capitol include o discutare mai pe larg a operei *Lucien Leuwen*. Ideea dezvoltărilor în aval, respectiv, în amonte este inspirată de D. Kahneman și A. Tversky, „The simulation heuristics", în D. Kahneman, P. Slovic și A. Tversky (editori), *Judgment Under Uncertainty* (Cambridge University Press, 1982). Interpretările la *Mansfield Park* pe care le critic aici le aparțin lui J. Heydt-Stevenson, „«Slipping into the ha-ha»: Bawdy humor and body politics in Jane Austen's novels", *Nineteenth-Century Literature*, 55 (2000), pp. 309-339, și lui J. Davidson, *Hypocrisy and the Politics of Politeness* (Cambridge University Press, 2004).

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

BCU IASI/CENTRAL UNIVERSITY LIBRARY

IV
LECȚII DESPRINSE DIN ȘTIINȚELE NATURII

Diferitele discipline științifice se pot afla într-una din două relații: *reducție* sau *analogie*. Reducția ia forma explicării fenomenelor la un nivel al ierarhiei științelor în termenii unor fenomene de nivel inferior (vezi Figura IV.1).

Programele reducționiste tind să fie foarte controversate. Pentru o perioadă lungă de timp, mulți biologi au susținut cu vehemență că reducerea biologiei la chimie nu poate funcționa – însă a funcționat. De la Emile Durkheim încoace, mulți au susținut că științele sociale nu pot fi reduse la psihologie. O teză centrală a lucrării de față este că pot fi reduse astfel.

Între aceste două reducții se află aceea a psihologiei la biologie. Disciplinele biologice relevante sunt: genetica, fiziologia, biologia dezvoltării și biologia evoluționistă. Primele două studiază cauzele proxime ale structurii și comportamentului organismelor, ultimele două studiază cauzele îndepărtate în istorie ale organismului individual sau ale speciei. Studiul structurii și studiul comportamentului sunt înrudite, prin

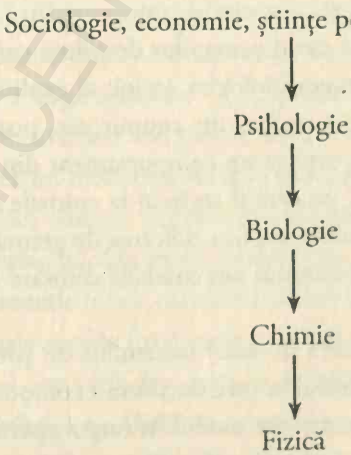

FIGURA IV.1

BCU IASI CENTRAL UNIVERSITY LIBRARY

aceea că structura oferă atât oportunități pentru comportament, cât și constrângeri asupra comportamentului. Faptul că avem doi rinichi, dar ne este necesar doar unul ne permite să-l donăm pe unul dintre ei unei rude pentru transplant și face posibile norme sociale care pot autoriza sau cenzura apariția respectivei practici. Însă motivul pentru care avem doi rinichi nu este acela de a permite transplantul de la donatori vii. Multe structuri există pentru că ne permit să facem anumite lucruri, dar acesta nu este un exemplu. De multe ori, însă, este dificil să ne dăm seama dacă permisivitatea unei structuri este accidentală sau explicativă.

Relevanța biologiei pentru celelalte științe sociale ar trebui să fie evidentă din moment ce domeniile lor se intersectează. Însă foarte mulți cercetători din științele sociale resping explicațiile biologice pe motiv că sunt „reducționiste". Aceasta este o acuzație stranie pentru aceia care cred, la fel ca mine, că reducționismul este motorul progresului în știință. Însă dacă „reducționismul" este însoțit de atribute precum „prematur", „brut" sau „speculativ", atunci obiecția poate fi întemeiată.

Reducționismul *prematur* se observă când oamenii de știință, convinși de fezabilitatea esențială a trecerii de la o explicație de nivel înalt la una de nivel inferior, încearcă să facă această trecere înainte să fie îndeplinite condițiile tehnice de măsurare, conceptele și teoriile aferente. Un exemplu clasic este fiziologia mecanicistă a lui Descartes, pe care Pascal o comenta în modul următor: „Descartes. Putem spune pe scurt: «Acest lucru este alcătuit din figură și mișcare», întrucât este adevărat. Dar, pentru a spune ce sunt toate acestea și pentru a alcătui mașinăria, este ridicol. Întrucât este inutil, nesigur și dureros." În ziua de azi, aceia care propun algoritmi pentru recunoașterea și clasificarea imaginilor sau pentru traducerea automată se pot găsi într-o situație asemănătoare. Sarcinile recunoașterii feței umane și a detectării propozițiilor fără sens pe care noi le executăm fără niciun efort, se află, cel puțin până în prezent, dincolo de capacitatea sistemelor artificiale.

Reducționismul *brut* se observă în cazul oamenilor de știință care încearcă să explice comportamente specifice în termeni biologici, în loc să explice *capacitatea* sau *tendința* unui comportament anume care, într-un anumit caz, poate sau nu să fie utilizată sau realizată. Încercarea de a explica un comportament din sfera politicului făcând apel la „imperativul teritorial" pe care îl regăsim la animale este un astfel de exemplu. Alt exemplu este ideea că putem explica ridicarea de greutăți ca un rezultat al selecției sexuale, analog cu penele păunului sau coarnele cerbilor. Se pot da multe alte exemple.

Reducționismul *speculativ* se observă în cazul oamenilor de știință care produc ipoteze fanteziste care dau seamă de modul în care un anumit comportament *ar fi putut* să apară, fără să arate și că acesta chiar *este* modul în care a apărut. Sociobiologia și domeniul înrudit al psihologiei evoluționiste abundă în astfel de exemple, ca, de exemplu, atunci când oamenii de știință susțin că autoamăgirea a evoluat pentru

beneficiile sale evoluționiste sau că depresia postnatală în cazul femeilor a evoluat un instrument de negociere (capitolul 17).

Nu ne ajută prea mult dacă spunem că reducționismul este ceva rău. Dorința de a reduce fenomene complexe la unele mai simple poate lua forme simpliste, însă acest lucru se poate întâmpla în cazul oricărei strategii de cercetare științifică. Istoria științei ne pune în fața unui număr covârșitor de cazuri în care reducționismul este progresist, iar antireducționismul acționează ca o forță reacționară în știință. De asemenea, istoria ne arată riscul pe care îl implică utilizarea *analogiilor* între o disciplină științifică și alta pentru generarea de ipoteze. În sine, utilizarea analogiilor este benignă: ipotezele științifice trebuie judecate după descendenții lor (implicațiile testabile), și nu după strămoșii lor. Însă, când gândirea prin analogie îi face pe oamenii de știință să privilegieze un an anumit tip de ipoteză în fața altora, rezultatul ajunge cel mai adesea în dulapul cu monstruozități al gândirii științifice. Analogia dintre societate și organismele biologice, de exemplu, a fost utilizată pentru a sprijini ideea că societățile, la fel ca organismele, sunt entități care se autoreglementează prin mecanisme de corecție, prin care echilibrul se menține în urma unor ajustări specifice (de exemplu, revoluțiile). În secolul al XIX-lea, oamenii de știință dezbăteau ce anume, în societate, ar corespunde celulei din organism, fără să se întrebe dacă există vreun motiv pentru a admite o astfel de analogie. Alți autori au utilizat analogiile fizice, mai degrabă decât pe acelea biologice, căutând echivalentul social al legilor forței gravitaționale ale lui Newton. Oamenii de știință care susțin că științele sociale pot avea un impact asupra obiectului pe care îl studiază invocă principiul incertitudinii al lui Heisenberg de parcă profunzimea acestui principiu ar putea să transforme un truism în ceva la fel de profund.

În capitolul 15 mă voi ocupa de anumite descoperiri din fiziologie și neuroștiințe care promit să ofere abordări reducționiste ale unor fenomene precum frica, încrederea și „graba de a trage concluzii". În capitolele 16 și 17 voi discuta teoria selecției naturale, care a fost utilizată atât pentru reducție, cât și pentru analogie. Susțin că în timp ce unele încercări reducționiste sunt plauzibile, altele sunt premature, brute sau speculative. Folosirea selecției naturale ca analogie pentru fenomenele sociale are o valoare extrem de dubioasă. Un motiv pentru care, în capitolul 16, voi explicita mecanismul selecției naturale într-o manieră care poate părea excesiv de detaliată pentru o carte despre științele sociale (însă mult prea superficială pentru a servi ca expunere a subiectului în sine) este acela de a arăta că, în lumea socială, nu există nimic cu care să se asemene, nici măcar vag. Vorbăriei unora dintre oamenii de știință despre „selecția socială" sau despre „evoluția socială" îi lipsește prea mult precizia și concentrarea pentru a fi luată în serios.

◈◈◈

Notă bibliografică

Pentru critici aduse sociobiologiei speculative și psihologiei evoluționiste, vezi P. Kitcher, *Vaulting Ambition* (Cambridge, MA, MIT Press, 1987) și D. Buller, *Adapting Minds* (Cambridge, MA, MIT Press, 2005). Pentru un argument referitor la faptul că abordarea prin neuroștiințe a comportamentului este prematură și pentru o apărare explicită a abordării de tip „cutie neagră", vezi J.R. Staddon, *Adaptive Dynamics: The Theoretical Explanation of Behavior* (Cambridge, MA, MIT Press, 2001). O trecere în revistă a metaforelor biologice în studiul societății se găsește în J. Schlanger, *Les métaphores de l'organisme* (Paris, Vrin, 1971).

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 15

FIZIOLOGIE ȘI NEUROȘTIINȚE

Căutarea bazelor fiziologice (de multe ori, neurofiziologice) ale comportamentului uman complex a primit un suflu nou în ultimii ani, în special ca rezultat al unor noi tehnici de măsurare și observație. Nu încape îndoială că o astfel de manieră de investigație are un viitor strălucitor, chiar dacă unele dintre încercările recente pot fi premature, brute sau speculative. Mă voi referi la trei seturi de descoperiri care par a fi în mod special relevante pentru scopurile cărții de față.

Frica

În capitolul 8 am susținut că emoțiile sunt declanșate în mod obișnuit de credințe sau de antecedente cognitive. Devin mânios dacă îmi imaginez că te-ai ciocnit de mine intenționat, din neglijență sau pentru că nu te uiți în jur, dar nu dacă aș crede că ciocnirea a fost rezultatul ciocnirii altcuiva de tine din cauza unei mișcări neașteptate a trenului. În *Prizoniera* lui Proust, sentimentele de gelozie ale naratorului cresc și descresc pe măsură ce se schimbă credințele sale referitoare la ce face Albertine în momentele în care nu o poate urmări. Însă ne putem întreba dacă nu cumva emoțiile nu sunt uneori cauzate de simple percepții care, spre deosebire de credințe, nu sunt „despre" ceva anume. De pildă, Descartes credea că surpriza sau uluirea „se pot petrece înainte să știm câtuși de puțin dacă respectivul obiect ni se potrivește sau nu".

Cercetări în neurofiziologie cu privire la frică (în cazul șobolanilor) confirmă această idee. Există două căi diferite care pornesc de la aparatul senzorial din talamus spre amigdala creierului (partea creierului care cauzează răspunsuri emoționale viscerale, precum și comportamentale). Confirmând viziunea tradițională că emoțiile sunt întotdeauna precedate și declanșate de un element cognitiv, o cale merge de la talamus la neocortex, partea creierului responsabilă cu gândirea, și de la neocortex mai departe, spre amigdală. Organismul primește un semnal, își formează o credință cu privire la ce înseamnă acest semnal, iar apoi reacționează emoțional. Există, însă, și o cale directă de la talamus spre amigdală, care ocolește în întregime partea gânditoare a creierului. În comparație cu prima cale, însă, cea de-a doua este „rapidă și improvizată". Pe

de o parte, este mai rapidă. În cazul unui șobolan, este nevoie de aproximativ 12 milisecunde (12 miimi dintr-o secundă) pentru ca un stimul acustic să ajungă de la amigdală prin calea talamică și aproape de două ori mai mult prin calea corticală. Pe de altă parte, cea de-a doua cale diferențiază mult mai puțin între diversele semnale primite. În timp ce cortexul poate să își dea seama că o formă curbată aflată în drum este doar un băț curbat, și nu un șarpe, amigdala nu poate face o astfel de distincție. Însă, din punctul de vedere al supraviețuirii, costul reacționării la vederea unui băț ca și cum ar fi vorba despre un șarpe trebuie să fi fost mult mai mic decât acela al celeilalte greșeli.

Nu se știe dacă aceste descoperiri din studiul fricii pot fi generalizate la alte emoții. Putem specula că și în cazul mâniei este adevărat ceva asemănător. Când suntem expuși la un eveniment care ar putea fi un atac, costul de oportunitate presupus de așteptarea de a vedea dacă într-adevăr este vorba despre o amenințare poate fi foarte mare. Selecția naturală poate foarte bine să fi implementat o tendință de tipul „trage mai întâi, pune întrebările mai încolo". Dacă reacționez și, mai târziu, aflu că nu eram victima unui atac, pot, totuși, să inventez o istorisire prin care să îmi justific comportamentul. Acest mecanism destul de subtil, legat de nevoia noastră de respect de sine și rezumat în proverbul „Cine comite o faptă rea nu poate ierta", ar putea interacționa cu un mecanism neurofiziologic pe care îl avem în comun cu animalele cărora le lipsește stima de sine. Sunt șanse mari ca multe dintre descoperirile din fiziologie și neuroștiințe să fie croite pe același calapod. Reacțiile aproape automate pe care le avem în comun cu alte specii pot fi supuse interpretărilor în propriul interes și dezvoltărilor specifice ființelor umane. Aceste raționalizări nu sunt banale din moment ce ne pot determina să persistăm în starea de agresiune, în loc să admitem că suntem de vină.

O altă descoperire din același program de cercetare sugerează o alternativă la teoria lui Freud despre reprimarea amintirilor. Victimele unui viol sau soldații care au ajuns pe câmpul de luptă nu își aduc aminte, uneori, nimic din experiențele prin care au trecut. Este neaducerea aminte (nevoia pentru acest termen neutru se va vedea în curând) un proces motivat, așa cum s-a gândit Freud, sau unul în care motivațiile nu joacă niciun rol? Răspunsul poate să depindă de relația dintre stres și formarea (sau non-formarea) amintirilor. Stresul ușor sporește amintirea unui eveniment stresant, dar stresul puternic și îndelungat poate ridica nivelul steroizilor generați de glandele suprarenale atât de mult, încât hipocampul (partea creierului unde se formează amintirile conștiente) este afectat în mod nefavorabil. Este posibil ca amintirile unor evenimente traumatizante să nu se fi format deloc și nu să fi fost reprimate. Dacă această concluzie este corectă, nu înseamnă că evenimentul traumatizant nu a lăsat nicio urmă psihică. Să presupunem că ești implicat într-un accident de mașină în care claxonul rămâne blocat. Ulterior, sunetul unui claxon devine un stimul condiționat al fricii care merge direct la amigdală și suscită răspunsuri organice care apar în mod

obișnuit doar în cazul existenței unui pericol. Emoția de panică, de asemenea, este declanșată de percepți, mai degrabă decât de o credință.

Încrederea

Voi anticipa discuția privitoare la încredere din capitolul 21 menționând trei descoperiri experimentale provocatoare. Toate se bazează pe un Joc al Încrederii (JÎ) între doi jucători, un „investitor" și un „administrator". Investitorul primește anumite fonduri, pe o parte dintre acestea putându-le transfera către administrator. Conducătorul experimentului fructifică apoi transferul, astfel încât administratorul primește de mai multe ori suma transferată. În cele din urmă, administratorul decide dacă trebuie să „returneze transferul" către investitor, iar, dacă o face, decide cât de generos urmează să fie. În anumite experimente, investitorul are și opțiunea de a-l pedepsi pe administrator dacă transferul returnat nu este suficient de generos. Dacă ambii jucători sunt raționali, își urmăresc propriile interese și știu fiecare despre celălalt că au aceste calități, atunci în situații unice și anonime nu va avea loc niciun transfer, însă, în experimente, niveluri pozitive de transferuri și de transferuri returnate constituie regula. Altă versiune a JÎ va fi pusă în discuție în capitolul 21. Aici voi menționa doar unele descoperiri care stabilesc o conexiune între comportamentul investitorilor, starea lor hormonală și activarea centrilor plăcerii din creier.

Primul experiment studiază dimensiunea investiției ca o funcție a prezenței sau absenței hormonului oxitocină. Se știe că hormonul stimulează comportamente prosociale în rândul rozătoarelor și stimulează lactația, dar descoperirea că stimulează comportamente prosociale sau încrederea în rândul oamenilor a venit ca o surpriză. Odată administrat hormonul, procentul investitorilor care și-au transferat întregul fond către administratori a crescut de la 21 la 45%. Alte trei descoperiri sunt foarte interesante. În primul rând, administratorii cărora li s-a administrat hormonul nu au returnat transferuri mai substanțiale. În al doilea rând, investitorii cărora li s-a administrat hormonul aveau aceleași credințe cu privire la încrederea în administratori (adică așteptări cu privire la transferurile returnate) ca și investitorii cărora hormonul nu le fusese administrat. În al treilea rând, când investitorii știau că returnarea transferurilor este generată de un mecanism întâmplător cu aceeași distribuție a recompenselor ca în cazul în care ar fi jucat împotriva unei persoane reale, oxitocina nu făcea nicio diferență pentru dimensiunea transferurilor.[1] O interpretare naturală este aceea că hormonul afectează comportamentul făcându-i pe investitori să aibă o „aversiune

[1] Recompensele reale ale unor investitori în JÎ împotriva unui administrator real au fost folosite pentru a determina distribuția recompenselor pe care li se spune că ar trebui să le aștepte atunci când joacă împotriva unui mecanism întâmplător. Presupunând că subiecții unui JÎ au așteptări raționale, adică au credințe corecte despre distribuția recompenselor pe care se pot

mai mică față de trădare", și nu reducându-le aversiunea față de risc. Importanța aversiunii față de trădare este confirmată și de alte experimente care nu se bazează pe manipulări fiziologice.

Cel de-al doilea experiment, care le permite investitorilor să îi pedepsească pe administratorii insuficient de generoși, urmărește să vadă ce se petrece în mintea acestora în timp ce execută pedepsele. În acest JÎ, investitorul are de ales între a transfera în întregime fondurilor sale în valoare de 10 unități monetare (UM) către administrator și a nu transfera nimic. Dacă face un transfer, acesta este mărit de patru ori de conducătorul experimentului, ajungând ca administratorul să aibă un total de 50 UM, adică cele 10 UM ale investitorului, plus cele 40 generate de investiție. Apoi administratorul are de ales între a returna 25 UM din 50 către investitor și a nu transfera nimic. Cu alte cuvinte, cele trei rezultate posibile sunt (10, 10), (25, 25) și (0, 50) (vezi Figura 15.1).

În plus, după ce administratorul ia o decizie, ambii jucători primesc un fond suplimentar de 20 UM. Investitorul poate să utilizeze fondul pentru a pedepsi administratorul în două moduri posibile. În modul „costisitor", investitorul poate acorda până la 20 de „puncte de pedeapsă" administratorului; fiecare punct determină o pierdere de 1 UM a investitorului și o pierdere de 2 UM a administratorului. Astfel, printr-o pedeapsă maximă, investitorul ar putea să se asigure că recompensa administratorului este redusă de la 70 (50 + 20) la 30, în timp ce recompensa lui este redusă de la 20 la 0.[1] Într-o variantă „fără costuri", doar administratorul este afectat de pedeapsă.

Din 15 investitori, toți cu excepția unuia au ales să facă transferuri. Experimentul a fost manipulat astfel încât fiecare investitor să joace împotriva a șapte administratori, dintre care trei returnează transferuri, în timp ce patru țin totul pentru ei. Acești administratori egoiști au format scopul experimentului. După ce un administrator își anunță decizia de a ține totul pentru sine, investitorul are la dispoziție un minut pentru a delibera și a decide dacă vrea să pedepsească administratorul și cât de severă va fi pedeapsa. În timpul acestei perioade, creierul său este scanat pentru a se detecta activități ale diverselor regiuni care ar putea fi relevante. Una dintre regiuni, nucleul caudat, este strâns legată de procesarea recompenselor. Altă regiune, cortexul prefrontal și cel

---

așteptă să le primească, aceștia sunt puși în fața aceleiași distribuții în ambele moduri. Această presupoziție poate fi pusă sub semnul întrebării.

[1] În sumele realmente plătite către cei supuși experimentului, fiecare punct corespunde sumei de 0,1 franci elvețieni. Suma pare foarte mică, dar alte experimente au arătat că, în jocuri de acest tip, se obțin aceleași rezultate din punct de vedere calitativ dacă se ridică miza până la un salariu lunar. Aceste rezultate sunt obținute în mare parte folosind burse de cercetare din Țările bogate pentru a desfășura experimente în țări din Lumea a Treia.

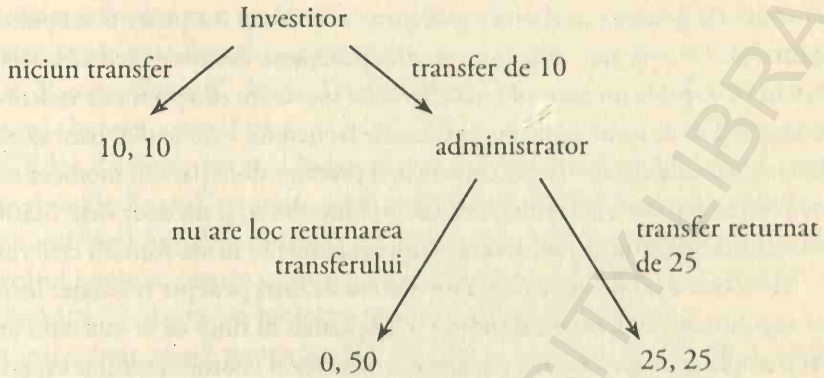

FIGURA 15.1

orbitofrontal, este legată de integrarea unor procese cognitive separate, de exemplu, a compensațiilor între costuri și beneficii. În fiecare dintre aceste regiuni, caracterul activităților care au loc confirmă ipoteza cu privire la motivația pedepselor, pe care o voi enunța în cele ce urmează.

Atât în modul costisitor, cât și în cel fără costuri există o corelație între activarea circuitelor legate de recompense și pedeapsa în bani impusă de fapt. Această corelație ar putea sugera fie că decizia de a pedepsi induce satisfacție, fie că satisfacția așteptată în urma pedepsei determină decizia de a pedepsi. Pentru a distinge între cele două ipoteze, conducătorii experimentelor au luat în considerare 11 subiecți care au impus pedeapsa maxim fezabilă în modul „fără costuri". Printre acești subiecți, cei în cazul cărora circuitele de recompensă s-au activat cel mai puternic au impus și cele mai stricte pedepse în modul „costisitor". Pe măsură ce pedeapsa le făcea mai multă plăcere, aceștia deveneau din ce în ce mai dispuși să cheltuiască mai mult pentru ea, susținând astfel cea de-a doua ipoteză. Această interpretare este confirmată, de asemenea, de faptul că cortexul se activa mai puternic în modul costisitor, când subiecții trebuiau să evalueze în contrapartidă costurile materiale și beneficiile psihice ale pedepsei, decât în modul fără costuri.

Această descoperire pare să confirme teoria „căldurii interioare" a acestei forme specifice de comportament altruist (capitolul 5). Pedeapsa, deși altruistă în acest mod comportamental indirect al beneficiilor aduse terților, implicând și anumite costuri pentru agent, nu este generată de motivații altruiste. Este prea devreme să ne dăm seama dacă rezultatele se vor menține pe viitor și dacă pot fi generalizate la formele directe de altruism (bani oferiți săracilor) sau la pedeapsirea terților (A îl pedepsește pe B pentru că l-a rănit în mod nedrept pe C). Imparțialitatea pedepsirii terților poate că nu produce satisfacția dulce care însoțește răzbunarea pentru un afront personal. Ce pare însă clar este că știința ne oferă acum mijloace prin care putem decide între ipoteze motivaționale concurente care eludează simpla introspecție. Dacă pedeapsa este

motivată de plăcerea așteptată a pedepsei, acest scop nu poate fi scopul conștient al agentului. Cu excepția sadicilor, oamenii pedepsesc deoarece cred că celălalt *o merită*. Putem face apel la un exemplu mai clar: este incoerent din punct de vedere conceptual să ne gândim că motivațiile conștiente ale oamenilor care le dau bani săracilor se pot datora *doar* căldurii interioare pe care le-o produce donația, din moment ce acel sentiment este un parazit al credinței că fac un bine altora, și nu doar sieși. Stările care sunt inaccesibile conștiinței pot deveni acum accesibile în urma scanării creierului.

Aplicarea unei pedepse diferă de *observarea* unei pedepse meritate. Într-un al treilea experiment, creierul unui subiect a fost scanat în timp ce se uita cum un șoc dureros era aplicat asupra a două persoane (complici ai coordonatorului experimentului), care interacționaseră cu subiectul într-un Joc al Încrederii, unul comportându-se corect, altul incorect. Când subiectul observă pedeapsa aplicată asupra jucătorului corect, scanarea creierului arată activarea unor zone legate de durere („Îți simt durerea"). Aplicarea pedepsei asupra jucătorului incorect a cauzat mai puțină activitate în zonele legate de durere, dar a crescut activitatea în zonele legate de plăcere (*Schadenfreude*). (O descoperire importantă, dar neexplicată, a fost aceea că femeile simțeau mai multă durere și mai puțină plăcere decât bărbații când erau pedepsiți subiecții incorecți.) *Schadenfreude* este un soi de mânie indirectă: A obține plăcere din a-l observa pe B cum îl pedepsește pe C, când C i-a făcut rău lui A. Ar fi interesant de investigat activitatea cerebrală în cazul indignării indirecte: oare A ar obține la fel de multă plăcere observând cum B îl pedepsește pe C, când C i-a făcut rău lui D? De asemenea, s-ar putea compara mânia indirectă cu mânia directă: în ce raport se află plăcerea obținută din observarea lui B pedepsindu-l pe C, când C i-a făcut rău lui A, cu plăcerea obținută din observarea lui A pedepsindu-l pe C, când C i-a făcut rău lui A? Aceste emoții elementare bine-cunoscute din literatură, piese de teatru și poezie par acum a fi inteligibile și pentru știință.

Completarea golurilor

Cititorii care folosesc corectoare automate de text știu că programele sunt capabile să „ghicească" sau să „completeze golurile" care ar putea să lipsească dintr-un cuvânt scris greșit. Uneori, aceste corectoare produc soluția greșită, care, dacă nu este verificată, poate avea efecte comice sau chiar dezastruoase, însă, în mare, aceste instrumente sunt de încredere și utile. Și creierul are această capacitate de a umple golurile lăsate libere de experiență. O figură circulară cu goluri mici va fi considerată completă, un cerc închis. Dacă există multe goluri, poate că nu va fi percepută literalmente ca un cerc închis, dar tot va fi perceput cercul invizibil din care fac parte elementele vizibile. În partea de sus a Figurii 15.2, „vedem" cercurile sub dreptunghiuri, dar când dreptunghiurile sunt îndepărtate, cum se întâmplă în partea de jos, dispar și cercurile.

Completarea golurilor este un fenomen cotidian. Deunăzi făceam o plimbare în Levallois-Peret, la periferia Parisului, și am remarcat un indicator stradal pe care se putea citi „Rue Auguste Blanqui". M-am întrebat de ce, într-un cartier atât de bogat și conservator, s-ar boteza o stradă cu numele celui mai violent revoluționar francez din secolul al XIX-lea. Uitându-mă mai îndeaproape, am observat că era, de fapt, vorba despre „Rue Aristide Briand", numele unui respectabil om de stat francez. Evident, am remarcat mai întâi inițialele numelui și prenumelui, precum și lungimea aproximativă a fiecărui nume și, pentru că mult mai familiar îmi este Blanqui decât Briand, am inferat în minte că observ un indicator stradal care îl comemorează.

Nu am interpretat aceste semne umplând în mod conștient literele lipsă: completarea s-a petrecut inconștient, iar experiența mea conștientă a fost aceea de a *vedea* două cuvinte complete. Cu altă ocazie, căutam o brutărie (*boulangerie*) și am observat un semn care indica un astfel de magazin, doar pentru a descoperi că observasem de fapt a doua parte din numele „Restaurant de l'Orangerie". Spre deosebire de situația

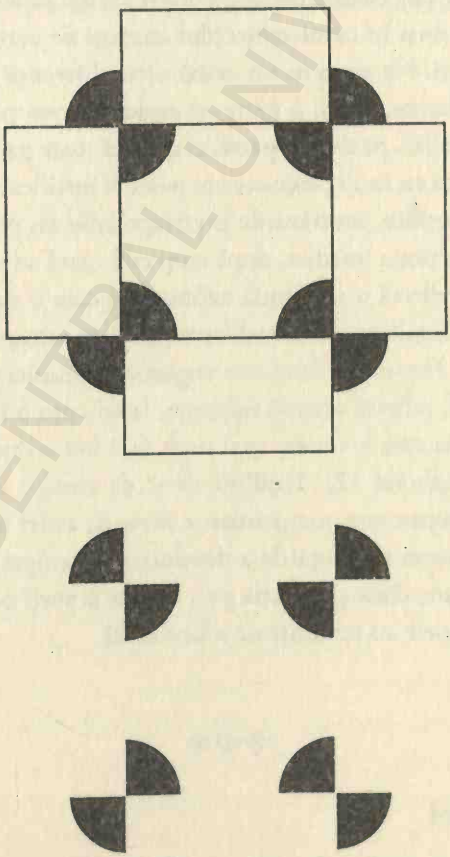

FIGURA 15.2

BCU IASI CENTRAL UNIVERSITY LIBRARY

Blanqui-Briand, în acest caz se poate să fi fost vorba despre *percepție greșită motivată*, și nu doar despre încercarea creierului de a „aprecia cât mai bine" ceea ce vede.

Oferirea celui mai bun răspuns se mai numește uneori și „a trage concluzii pripite". Creierul face acest lucru tot timpul, din fericire pentru noi. Mecanismele de completare a spațiilor goale, precum corectoarele de text, sunt utile, ba chiar indispensabile. Totuși, uneori căutarea unor trăsături comune poate deveni o căutare obsesivă a sensului. În mod natural, creierul vede peste tot teorii ale conspirației. Mai precis, conform unor anumite teorii, emisfera dreaptă a creierului are funcția de a impune un cadru coerent asupra tuturor informațiilor cu care suntem bombardați în permanență. Se pot explica astfel unele cazuri de „anosognozie" sau de negare a bolii de către bolnavi. Mai degrabă decât unui fenomen *motivat*, cum ar sugera abordarea freudiană, negarea i se poate datora pur și simplu celui mai bun răspuns oferit de corectorul automat din creier. Se pare că fenomenul se datorează atât unor mecanisme motivaționale, cât și unor mecanisme nemotivaționale.

Cred că activitățile de „completare a golurilor" și de „acordare de sens" ale creierului funcționează nu doar în cazul subiecților studiați de oamenii de știință, ci și în cazul multor cercetători. Nu spun că nu există nicio diferență între teoriile conspiraționiste ale grupării Hamas, pe de o parte, și critica literară postmodernă, explicația funcțională sau psihanaliză, pe de altă parte, ci doar că toate par a fi animate de nevoia de a găsi sens și coerență cu mult peste ceea ce poate fi justificat de faptele disponibile. În cazul ipotezelor științifice, scurtăturile și extrapolările nu pot fi justificate și explicate prin nevoia de a acționa imediat, după cum este cazul adesea în viața de zi cu zi. Mai degrabă, acestea reflectă o profundă neliniște pe care o simțim când nu suntem capabili să-i atribuim lumii un sens, indiferent dacă se așteaptă sau nu să acționăm asupra ei în vreun fel. Nevoia de finalitate cognitivă și intoleranța față de acceptarea ignoranței (capitolul 7) reflectă această neliniște, la fel cum o face și hiperraționalitatea manifestată de aceia care investesc mai mult în a lua o decizie rațională decât în miza deciziei înseși (capitolul 12). Tendința de a „da vina pe victimă" este alt exemplu. Presupunând că lumea este esențialmente dreaptă, astfel încât oamenii „primesc ceea ce li se cuvine", avem tendința de a devaloriza și denigra victimele unor evenimente pur întâmplătoare, chiar și selecția prin tragere la sorți pentru serviciul militar; într-adevăr, chiar victimele au tendința de a face la fel.

◈◈◈

Notă bibliografică

Discutarea fricii se bazează pe J. LeDoux, *The Emotional Brain* (New York, Simon & Schuster, 1996). Impactul oxitocinei asupra încrederii este discutat în M. Kostfeld

et al., „Oxytocin increases trust in humans", *Nature*, 435 (2005), pp. 673-676. Ideea aversiunii față de trădare este confirmată de I. Bohnet și R. Zeckhauser, „Trust, risk and betrayal", *Journal of Economic Behavior and Organization*, 55 (2004), pp. 467-484. Studiul încrederii și răzbunării îi aparține lui J.F. de Quervain *et al.*, „The neural basis of altruistic punishment", *Science*, 305 (2004), pp. 1254-1258. Studiul încrederii și al conceptului *Schadenfreude* îi aparține lui T. Singer *et al.*, „Empathic neutral responses are modulated by the perceived fairness of others", *Nature*, 439 (2006), pp. 466-469. Un ghid interesant al fenomenului de „umplere de goluri" îi aparține lui V.S. Ramachandran și S. Blakeslee, *Phantoms in the Brain* (New York, Quill, 1998), care explică cum ostilitatea inițială a lui Ramachandran față de explicațiile freudiene ale anosognoziei au fost înlocuite de convingerea că un element de negare inconștientă trebuie să fie prezent; vezi și N. Levy, „Self-deception without thought experiments", în au.geocities.com/neil_levy/Documents/articles/Self-deception.pdf. Descoperirile privitoare la „învinovățirea victimei" se găsesc în S. Rubin și A. Pepau, „Belief in a just world and reaction to another's lot", *Journal of Social Issues*, 29 (1973), pp. 73-93.

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

Capitolul 16

EXPLICAȚIA PRIN CONSECINȚE ȘI SELECȚIA NATURALĂ

Consolidarea

Toate explicațiile sunt explicații cauzale. Explicăm un eveniment precizându-i cauza. Cauzele își precedă efectele în timp. Decurge de aici că nu putem explica un eveniment, cum ar fi o acțiune, făcând apel la consecințele sale. Dacă, însă, explanandumul este un model al unui comportament recurent, consecințele acelui comportament, într-o anumită ocazie, pot fi parte a cauzelor care îi sporesc șansele de apariție într-o ocazie ulterioară. Există două moduri prin care se poate întâmpla așa ceva: prin *consolidare* sau prin *selecție*. Mă voi concentra pe al doilea mod din moment ce este mai important pentru scopurile mele, dar voi începe prin a spune câteva cuvinte despre primul mod.

În cazul în care consecințele unui anumit comportament sunt plăcute sau aduc recompense, avem tendința să adoptăm mai des acel comportament; în cazul în care sunt neplăcute sau aduc pedepse, comportamentul respectiv va fi mai rar. Mecanismul subiacent ar putea fi alegerea rațională conștientă, dacă observăm consecințele plăcute sau neplăcute și ne decidem să acționăm în viitor astfel încât să repetăm sau să evităm experiența respectivă.[1] Însă, de cele mai multe ori, consolidarea se poate petrece fără o alegere intenționată. Când bebelușii învață să plângă pentru că părinții îi recompensează luându-i în brațe de fiecare dată când o fac, nu avem motive să credem că, mai întâi, copiii observă beneficiile pe care le aduce plânsul și pe urmă plâng pentru a le obține. Când copiii mai mari fac scandal pentru a obține ceva, părinții își dau de obicei seama că nu au de a face cu o emoție autentică.

Învățarea prin consolidare a fost studiată intens în experimente de laborator pe animale. Într-un experiment tipic, animalului i se dă posibilitatea să apese o manetă sau una dintre mai multe manete, iar recompensa se acordă fie ca o funcție a

[1] Să ne aducem, însă, aminte că nu suntem întotdeauna în măsură să precizăm care dintre cele două experiențe este mai dureroasă (capitolul 12).

BCU CLUJ CENTRAL UNIVERSITY LIBRARY

numărului de apăsări ale manetei din momentul ultimei recompense, fie ca o funcție de timp din momentul ultimei recompense. În fiecare caz, funcția poate fi deterministă sau probabilistă. Într-un *plan cu proporție fixă*, animalul primește o recompensă după ce a apăsat pe o manetă de un anumit număr de ori, în timp ce într-un *plan cu proporție variabilă* numărul de apăsări de care este nevoie pentru a produce o recompensă variază în mod întâmplător. În fiecare caz, fiecare apăsare produce un „punct de recompensă" care se adaugă punctelor anterioare. În *planurile cu interval fix* o apăsare va produce o recompensă la un anumit interval de timp după ultima recompensă, în timp ce într-un *plan cu interval variabil* timpul de așteptare până la o nouă recompensă variază aleatoriu. În fiecare caz, momentul acordării recompensei este independent de numărul de apăsări. Fiecare plan de consolidare produce, după o anumită perioadă de învățare, un model comportamental specific și stabil care, mai mult, se va dizolva într-un model specific odată cu îndepărtarea factorului de consolidare (recompensa). De exemplu, reacțiile învățate prin recompensarea fiecărei apăsări de manetă (un caz special al planului cu proporție fixă cunoscut drept consolidare continuă) se dizolvă mai repede decât acelea învățate într-un plan întâmplător cu proporție variabilă. Intuiția ar părea să indice exact pe dos, din moment ce consolidarea continuă ar trebui să producă o deprindere mai puternică, dar, după cum se întâmplă uneori, intuiția se înșală.

Relevanța acestor descoperiri în afara cazurilor de laborator depinde de scop. Dacă scopul este acela de a *modela* acțiunea, de exemplu, într-o situație apărută în sala de curs, într-un cazinou sau la locul de muncă, un proiectant poate (mai mult sau mai puțin constrâns) să impună un plan de recompense care să genereze comportamentul dorit. De pildă, planurile cu interval variabil sunt adeseori utilizate pentru a modela comportamentul, ca atunci când un profesor anunță o strategie de teste întâmplătoare. În planul cu proporție variabilă, care funcționează în multe jocuri de noroc, este mai ușor de stabilit un comportament dacă prima recompensă apare devreme.[1] Întrucât managerii de cazinouri și de curse de cai nu au la dispoziție tehnologia pentru a atrage novici cu oferte ale unor câștiguri mari, aceștia trebuie să se bazeze pe faptul că, prin legile probabilității, unii pariori vor avea norocul începătorului.[2] Operațiunile de extorcare, însă, se bazează adesea pe stimularea deliberată cu ajutorul unor câștiguri timpurii. În sala de curs și în cazinou, planurile de recompense funcționează „pe la spatele" studenților sau al jucătorilor de noroc, în sensul că aceștia nu își modelează comportamentul prin stimulente explicite, ci, mai degrabă, ca în cazul bebelușului

[1] Este mult mai ușor de stabilit dacă tehnologia jocurilor de noroc îngăduie posibilitatea unor situații foarte apropiate de câștig. Cu toate că fiecare astfel de situație este mai puțin consolidantă decât un câștig concret, aceste situații sunt mai frecvente.

[2] Norocul lor este, în acest caz, ghinionul lor și norocul cazinoului.

care plânge, printr-un proces inconștient. Dimpotrivă, când directorii își plătesc angajații după ce au atins o țintă stabilită (un plan cu proporție fixă) sau pe o bază lunară (un plan cu interval fix), aceștia inițiază un sistem de recompense. Din moment ce comportamentul angajaților poate fi explicat în mod adecvat prin recurs la recompensele *așteptate*, nu trebuie să recurgem la recompensele *efective*.

Dacă scopul este acela de a explica modele comportamentale prin consecințele lor efective, planurile de recompensă sunt relevante doar dacă apar în mod natural și, de asemenea, dacă sunt atât de opace încât nu creează stimulente explicite. Așa ceva nu se întâmplă prea des cu cele două planuri fixe. În viața de zi cu zi, simplul număr de răspunsuri este rareori decisiv pentru acordarea unei recompense. Comportamentul prietenilor mei față de mine nu este modelat de numărul de zâmbete prietenoase pe care le afișez, ci de consistența și de caracterul corespunzător al acelor zâmbete în situația dată. În scenarii obișnuite, recompensele care apar din când în când, cum ar fi salariul, sunt rare. Cele două planuri variabile sunt mult mai importante. O persoană care joacă „rece și cald" (un plan cu proporție variabilă) cu un membru de sex opus (sau de același sex) poate determina o atracție mai puternică decât cineva care afișează permanent un comportament prietenos. Un plan cu interval variabil apare când încerci să dai de cineva la telefon, însă numărul este ocupat. Știi că, mai devreme sau mai târziu, vei reuși să te conectezi formând numărul în mod repetat, însă nu știi precis când anume. Această situație determină un model de formare constantă a numărului, care probabil că nu constituie predicția unică a teoriei alegerii raționale. Această teorie ar putea prezice oricâte modele, în funcție de credințele celui care dă telefon despre cât de mult va dura conversația cu celălalt. Pare, însă, greu de crezut că oamenii au convingeri stabile cu privire la astfel de chestiuni.

Prototipul de reacție generat de consolidare nu este, în general, acela produs de alegerea rațională conștientă. Să presupunem, de pildă, că un animal are de ales între apăsarea uneia dintre două manete, una recompensând conform unui plan cu proporție variabilă, iar cealaltă recompensând conform unui plan cu interval variabil. Modelul rațional care va maximiza recompensa globală este de a apăsa cel mai des maneta cu proporție variabilă pentru a acumula puncte de recompensă, vizitând din când în când și maneta cu interval variabil pentru a vedea dacă nu cumva este disponibilă vreo recompensă. Însă nu acesta este modelul produs de învățarea prin consolidare. Animalele nu fac nimic altceva decât să apese maneta cu interval variabil mult mai mult decât este optim. Făcând astfel, acestea *egalizează* recompensele medii ale apăsării uneia dintre manete, și nu recompensele *marginale*, așa cum ar dicta raționalitatea. Pentru alte combinații de planuri, învățarea prin consolidare imită uneori alegerea rațională, dar nu într-un mod consecvent. Dacă există vreun mecanism neintenționat

capabil să simuleze în mod temeinic raționalitatea, atunci va trebui să îl căutăm în altă parte.

Adecvarea reproductivă diferențiată

Cel mai des menționat mecanism este acela al selecției naturale sau sociale. În Partea a III-a a cărții am discutat cum putem explica un comportament presupunând că agentul se adaptează mediului într-un mod mai mult sau mai puțin rațional. Într-o perspectivă radical diferită, putem încerca să explicăm comportamentul admițând că *agenții sunt selectați de mediu.* Cu toate că selecția poate fi munca unui agent intențional, ca în cazul câinilor domestici care sunt dresați să fie docili sau în cazul șobolanilor de laborator care sunt dresați să fie mai inteligenți, multe mecanisme de selecție se bazează pe procese cauzale care nu presupun niciun agent intențional.

În special, *supraviețuirea diferențiată* a organismelor bazată pe modelele lor comportamentale poate conduce la un comportament optim (optim pentru reproducere) la nivelul populației, chiar în absența oricăror alegeri de optimizare sau intenții. Să presupunem că 10% dintre organismele unei populații de 100 de organisme se hrănesc atât de eficient, încât lasă în medie 10 urmași care ajung la vârsta maturității, pe când restul de 90% lasă doar 5. În cazul în care comportamentul părinților (nu contează prin ce mecanism) se transmite urmașilor, generația următoare de organisme adulte va include o proporție de 100/550 = 2/11, adică aproximativ 18% care manifestă comportamentul mai eficient. În cursul câtorva generații, aproape toate organismele vor manifesta același comportament. Dacă ne întrebăm de ce este manifestat la nivel global, răspunsul este: pentru că are consecințe mai bune.[1] Mecanismul funcționează de-a lungul mai multor generații. Spre deosebire de învățarea prin consolidare, nu modifică vreun comportament la nivelul unui anume individ, ci doar comportamentul tipic al unor generații succesive de indivizi.

Selecția naturală

Teoria selecției naturale exprimă această ipoteză în amănunt. Voi prezenta, în cele ce urmează, o versiune simplificată sau „clasică" a teoriei, care este suficientă pentru scopurile mele. Vreau să arăt, în special, modul în care selecția naturală diferă de alegerea intenționată și, cu atât mai mult, de alegerea rațională. Selecția naturală este un mecanism de optimizare, însă doar într-un sens slab. Vreau, de asemenea, să pun bazele pentru argumentul pe care urmează să îl ofer în capitolul următor, conform căruia selecția *socială* are șanse mici să genereze chiar și această formă slabă de optimizare.

---

[1] Mai bun în raport cu ce? Ne vom ocupa de această chestiune în curând.

Adecvarea unui organism, măsurată după numărul de urmași (rezultați prin reproducere)[1], este determinată atât de mediul acestuia, cât și de proprietățile fiziologice și comportamentale, adică de *fenotip*. *Genotipul* său este setul de instrucțiuni care, împreună cu mediul, determină fenotipul. Aceste instrucțiuni sunt codificate în moleculele lungi de ADN, care au multe proprietăți în comun cu un limbaj scris. Literele moleculei de ADN sunt patru molecule (nucleotide) numite T, A, G și C. Cuvintele sau codonii ADN sunt triplete de nucleotide, fiecare tripletă conținând instrucțiuni pentru constituirea unuia dintre cei 20 de aminoacizi, care constituie cărămizile de bază ale proteinelor din organism. Din moment ce există 64 de triplete (61, dacă le excludem pe cele trei cu funcție diferită), mai mult de una poate fi responsabilă pentru codificarea aceluiași aminoacid. O *genă* este un segment de ADN care codifică o proteină anume.

Organismele care se hrănesc mai eficient au apărut inițial ca rezultat al unei mutații întâmplătoare la nivelul genomului. Viziunea clasică a evoluției susține că mutațiile ar trebui să rezulte dintr-un număr mic de erori întâmplătoare în duplicarea materialului genetic, care are loc în procesul de reproducere. Există mai multe tipuri de mutații mici: ștergerea unei nucleotide, inserția unei nucleotide sau substituția unei nucleotide cu alta.

În cadrul acestei expuneri simplificate mă voi referi doar la cel de-al treilea tip de eroare, care este și cel mai des întâlnit. Cu toate că acum se știe cum pot mecanisme precum duplicarea genelor să producă mutații la nivel macro, voi rămâne la nivelul viziunii clasice pentru moment.

În cadrul acestei viziuni, modul în care mutațiile contribuie la evoluție poate fi clarificat prin recurgerea la o analogie cu structura pe trei niveluri a unei propoziții tipărite. Figura 16.1 ne arată asemănările dintre cele două cazuri.

Când o carte este culeasă din nou după prima ediție, neatenția culegătorului de text poate provoca anumite devieri de la textul original. Ne putem gândi că acestea presupun înlocuirea unei litere cu alta, ceea ce transformă cuvântul „mână" în „până" sau „nână". Erorile sunt întâmplătoare în sensul că nu au legătură cu conținutul cărții. În mod asemănător, mutațiile sunt întâmplătoare în sensul că șansa ca ele să apară nu are legătură cu natura modificărilor fenotipice pe care le vor determina. Anumite litere pot fi schimbate mai frecvent decât altele, în cazul în care culegătorul are dificultăți în a diferenția între „m" și „n". Ediția a doua a unor cărți poate conține mai multe

[1] Adecvarea în sens reproductiv este diferită de adecvarea ecologică, măsurată prin durata de viață. Cu toate că adecvarea ecologică este, de obicei, un mijloc pentru adecvarea reproductivă, cele două pot fi destul de divergente, cum este cazul părinților care își riscă propria viață pentru supraviețuirea urmașilor.

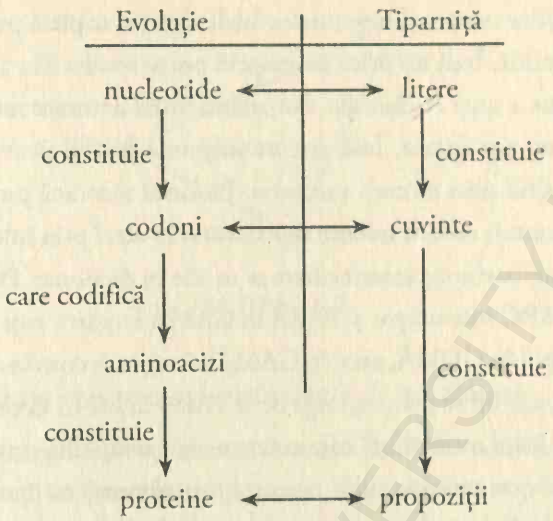

FIGURA 16.1

erori decât altele dacă, de exemplu, culegătorul era beat. Similar, ratele de mutație pot fi afectate de factori mutageni.[1]

Greșelile de tipar dintr-o carte sau dintr-un mesaj scris au, de obicei, un impact negativ asupra conținutului, uneori impactul fiind (literalmente) fatal. Deplasarea unei virgule, de exemplu, poate fi o chestiune de viață și de moarte.[2] De asemenea, cele mai multe mutații au consecințe dăunătoare pentru organismele la nivelul cărora apar, ceea ce înseamnă că adecvarea reproductivă scade. Din când în când, însă, poate apărea o mutație care sporește adecvarea unui organism, la fel cum eșecul accidental de a reproduce un text al primei ediții poate furniza informații mai corecte sau actualizate. În pasul următor al argumentului, însă, analogia care a funcționat până acum cedează. O mutație favorabilă va fi mult mai reprezentată la nivelul generației următoare din moment ce organismul în cadrul căruia a apărut va avea mai mulți urmași, iar mutațiile se transmit urmașilor. Dimpotrivă, nu există niciun mecanism prin care o carte în care greșelile din prima ediție s-au corectat accidental se va vinde mai bine.

În greșelile de tipar, orice literă poate fi înlocuită de alta. Cuvântul nou nu trebuie să aibă sens; cuvântul „mână" poate fi înlocuit la fel de bine cu „până" sau cu „nână".

---

[1] De pildă, radiațiile pot produce mutații; de aceea, scanarea cu raze X pentru detectarea cancerului poate cauza cancer prin determinarea unor mutații.

[2] Din moment ce nu am un echivalent în limba engleză, voi ilustra această afirmație recurgând la două propoziții în norvegiană: „Vent, ikke heng ham" și „Vent ikke, heng ham", care înseamnă „Așteaptă, nu-l spânzura" și, respectiv, „Nu mai aștepta, spânzură-l".

Și în cazul mutațiilor oricare dintre nucleotidele dintr-o tripletă poate fi înlocuită de oricare altă nucelotidă. Însă nu orice aminoacid poate rezulta din alt aminoacid printr-o simplă mutație a unei nucleotide. Pot apărea astfel anumite modificări ale proteinelor, dar altele nu pot apărea. Însă este întotdeauna posibil să se facă trecerea de la un aminoacid la altul prin mutații succesive. Procesul seamănă puțin cu jocul de familie în care un anumit cuvânt trebuie transformat în altul prin înlocuirea unei litere, cu condiția ca toate cuvintele intermediare să se afle în dicționar. Poți, astfel, să transformi PARĂ în PARC într-un pas și PARĂ în GAMĂ în patru pași (PARĂ în PURĂ, apoi în PUMĂ, apoi în GUMĂ, apoi în GAMĂ), însă sunt convins că, oricât de mare ar fi numărul de pași, nu se poate ajunge de la PARĂ la LINX decât dacă ieși în afara dicționarului. În „jocul mutațiilor" care corespunde exemplului, toți cei 64 de codoni sunt „în dicționar", cu excepția celor trei care funcționează ca „punct" (întrucât delimitează proteina). Cu toate că acești „codoni-stop" blochează modificările tripletei dintr-un pas, constrângerea e prea slabă pentru a bloca modificări în mai mulți pași. Astfel se poate ajunge de la aminoacidul metionină la triptofan în doi pași, fie de la ATG la TTG (leucină) la TGG, fie de la ATG la AGG (glicină) la TGG.[1] În plus, există căi care presupun mai mult de doi pași, precum ATG-ACG-TCG-TGG.

Ne putem imagina acum un joc de familie ceva mai complicat decât acela jucat în mod obișnuit; să presupunem că fiecare cuvânt este inclus într-o propoziție, iar cuvântul înlocuit nu trebuie doar să se afle în dicționar, ci trebuie să producă și o propoziție cu sens.[2] Există două constrângeri cu privire la înlocuirile de litere. Să presupunem că propoziția inițială este „Aduc prăjituri de casă". Înlocuind-o cu „Aduc prăjituri de acasă", respectăm atât constrângerea de sens, cât și pe aceea de dicționar. Înlocuind-o cu „Adun prăjituri de casă", respectăm doar constrângerea de dicționar. Înlocuind-o cu „Adun prăjituri de rasă" nu respectăm niciuna dintre condiții. Dacă „Adun prăjituri de casă" este modificată mai departe în „Adun prăjituri de masă", satisfacem ambele condiții. Pentru a ajunge la cea din urmă propoziție (cu sens) prin modificări aduse propoziției inițiale (cu sens) a trebuit să ieșim din setul propozițiilor cu sens.

---

[1] Am ales acest exemplu pentru că acești doi aminoacizi sunt singurii care sunt codificați de o singură tripletă. Toți ceilalți acizi au doi sau mai mulți codoni. Astfel, în contextul de față ar trebui să citim expresia „o cale de la aminoacidul X la aminoacidul Y" drept „o cale de la un codon dat pentru X la unul dintre codonii pentru Y".

[2] Acest tip de joc de limbaj era practicat de un grup de scriitori francezi cunoscut sub denumirea de „mișcarea Oulipo". Cel mai cunoscut membru al mișcării, Georges Perec, a scris un întreg roman, *La Disparition*, în care litera „e" nu este utilizată deloc.

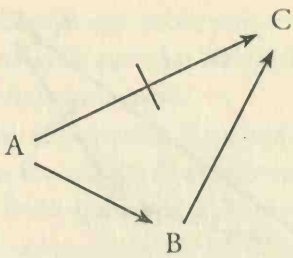

FIGURA 16.2

Selecția naturală generează puncte locale de maximum

Analogia cu biologia este evidentă. Un aminoacid care rezultă prin substituirea unei nucleotide este încorporat într-o proteină. Din moment ce proteinele sunt vitale pentru organism, ele trebuie să fie viabile din punct de vedere biologic (analog cu sensul pentru propoziții). Dacă oricare proteină intermediară nu este viabilă în drumul său de la forma inițială la cea finală, ea nu se poate fixa la nivelul populației *întrucât organismul în care a apărut pentru prima dată mutația spre acea proteină nu poate lăsa descendenți care să facă pasul următor*. Chiar dacă forma finală a proteinei ar crește adecvarea prin comparație cu cea din primul stadiu, acest fapt nu poate genera presiune evolutivă în stadiile intermediare.

Dacă *toate* căile fezabile de la primul până la cel din urmă stadiu ar condiționa organismul să adopte strategia indirectă a „unui pas înapoi, doi pași înainte", forma finală nu s-ar atinge niciodată. Din moment ce majoritatea și, probabil, vasta majoritate a mutațiilor sunt dăunătoare, această situație este destul de plauzibilă. Într-o formă simplificată (care nu ține cont de căile multiple), structura este reprodusă în Figura 16.2. Trei proteine diferite conferă grade diferite de adecvare (măsurate vertical). Structura codului genetic îngăduie mutații ale unei singure nucleotide de la A la B și de la B la C, dar nu de la A la C.

Selecția naturală (în această versiune clasică) este constrânsă la îmbunătățiri incrementale mici. Organismul urcă de-a lungul unei pante a adecvării până ajunge la un punct de *maximum local*, definit ca o stare în care orice altă modificare într-un singur pas ar reduce adecvarea. Cu toate că putem vorbi despre vârfuri mai înalte în „peisajul adaptiv", acestea nu pot fi atinse prin modificări într-un pas. Acest proces este diferit de alegerea intenționată în trei moduri. Să ne aducem aminte din capitolul 6 că, în virtutea intenționalității lor, ființele umane sunt capabile (1) să folosească strategii indirecte, (2) să aștepte și (3) să ochească o țintă în mișcare. Tocmai am văzut că selecția naturală nu este capabilă să facă pasul (1). Cât privește pasul (2), să luăm în considerare Figura 16.3.

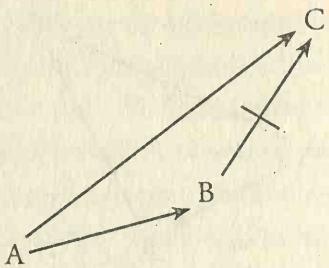

FIGURA 16.3

Structura codului genetic permite mutații ale unei singure nucleotide de la A la B și de la A la C, dar nu de la B la C. Dacă apare o mutație de la A la B, populația este prinsă în punctul de maximum (de nivel jos) B, din moment ce o mutație într-un pas la punctul global de maximum C este blocată. Nu există vreun mecanism care să favorizeze o mutație spre C mai degrabă decât spre B. Spre deosebire de agenții cu intenție, selecția naturală este *oportunistă*, ea profită de orice ocazie pentru îmbunătățiri, fiind în același timp incapabilă de a *aștepta* până la apariția unei modificări optime.

Cât privește pasul (3), populațiile se adaptează la un mediu care se află în schimbare permanentă. Dacă schimbările au loc cu regularitate, de pildă, după anotimpuri sau zilnic, acestea se adaptează la schimbări. Dacă apare un eveniment unic, cum ar fi o schimbare subită a climei, comportamentul care, înainte de schimbare, era la un punct adaptiv de maximum local poate deveni suboptim, astfel încât mutațiile care mai înainte ar fi fost dăunătoare sunt acum favorabile. Dacă schimbarea se prelungește, ca în cazul în care clima se răcește sau se încălzește de-a lungul unei perioade de timp, acest proces poate să nu ajungă niciodată la un punct de maximum local. Populația se va sincroniza cu schimbările din mediu cu o eficiență care depinde de viteza relativă a celor două procese. Adaptările uimitor de rafinate observate la plante și animale sugerează că animalele se adaptează la mediu mult mai rapid decât se schimbă mediul. Însă organismele vor rămâne întotdeauna puțin în urmă din moment ce nu pot *anticipa* schimbările de mediu. Dimpotrivă, ființele umane pot deveni conștiente de modificări viitoare, cum ar fi încălzirea globală, și pot să își ia precauții înainte ca acestea să se producă efectiv sau, dacă schimbările sunt rezultatul comportamentului uman, le pot împiedica apariția.

Mediul unei populații este alcătuit, printre altele, din populații aparținând altor specii, în raport cu care se poate afla într-o relație de pradă-prădător. Ca pradă, poate să dezvolte strategii de apărare mai bune; ca prădător, strategii de vânătoare mai bune. La fel cum o vulpe anume aleargă iepurele pe câmp, tot așa și specia Vulpe aleargă specia Iepure de-a lungul generațiilor. Însă, în timp ce logica selecției naturale împiedică Vulpea să anticipeze unde urmează să fie Iepurele peste câteva milenii, unii prădători

sunt capabili să intersecteze calea pe care prada vrea să fugă. Asemănător, procesul de maximizare locală a selecției naturale a produs capacitatea pentru maximizare globală, care poate fi regăsită în cazul ființelor umane.

Mutațiile sunt și un input, și un produs al selecției naturale, din moment ce rata la care acestea se produc poate fi modelată de efectele asupra adecvării. Mai multe mecanisme sunt implicate aici. Într-o populație de bacterii expusă constant la noi antibiotice, descendenții cu rate mai mari ale mutațiilor pot fi favorizați. Acesta este un caz special al ideii mai generale că, într-un mediu în schimbare, populațiile „imuabile" vor ajunge să dispară. În același timp, rate mari ale mutațiilor pot duce la pierderea unor informații genetice valoroase și, astfel, la dispariție. Undeva între aceste două extreme se găsește rata optimă. Într-un mediu constant, există o compensație între beneficiile unei rate mici a mutațiilor și costul „corectării" și reparării mecanismelor de care este nevoie pentru a menține o rată mică.

Această „viziune clasică" este simplistă din mai multe puncte de vedere, care ar trebui abordate dacă această carte ar fi un manual de biologie. Din moment ce lucrurile nu stau astfel, voi preciza doar că versiunea mai complicată modifică, dar nu invalidează, teza conform căreia selecția naturală are tendința să rămână la nivelul unei „capcane de maximum local". Mutații mari apar, iar unele dintre ele sunt responsabile pentru dezvoltări care, altminteri, nu ar fi fost posibile prin mutații mici. De asemenea, formele inferioare nu sunt eliminate imediat de concurență. În Figura 16.2, mutația spre B nu produce cu necesitate un organism neviabil în sensul strict al faptului de a fi incapabil să supraviețuiască sau să se reproducă. Unele organisme în starea B ar putea supraviețui pentru a produce organisme în starea C. În Figura 16.3, anumite organisme din starea A ar putea supraviețui competiției din partea organismelor mai eficiente din starea B îndeajuns de mult pentru ca mutația spre starea C să apară. Dacă maximumul global apare, aceasta este o chestiune care ține de viteza relativă a celor două procese: dispariția varietăților inferioare și rata de apariție a mutațiilor favorabile. Nu există, însă, vreun mecanism care ar putea să imite, *în mod sistematic*, capacitatea ființelor intenționale de a înlocui, de a aștepta sau de a folosi strategii indirecte.

Ca urmare a tendinței selecției naturale de a produce comportamente optimizante sau structuri optimale prin urcarea pantei, nu ar trebui să tragem concluzia că toate trăsăturile observate ale organismelor sunt optime. Pe de o parte, după cum am precizat deja, ar putea exista o *întârziere*, dacă populația nu a reușit încă să ajungă din urmă schimbările mediului. Ceea ce observăm poate fi optim pentru un anumit moment din trecut, dar nu pentru prezent. Pe de altă parte, fenomenul *pleiotropiei* ne spune că selecția naturală poate favoriza trăsături care sunt suboptime considerate izolat, însă, împreună cu alte trăsături, pot fi parte a unei soluții optime. Așa ceva este posibil pentru că o genă poate avea mai mult de un efect asupra fenotipului. Dacă efectele pozitive au mai multă greutate decât efectele negative, gena va fi favorizată de

selecția naturală. Ca exemplu, să ne gândim la existența unor trăsături masculine la insecte care provoacă vătămări fizice partenerilor în timpul actului sexual. S-a sugerat că aceste vătămări apar fie ca un efect secundar pleiotropic negativ al adaptării, care le dă masculilor un avantaj reproductiv într-un alt context, fie ca o adaptare a masculilor în general. Experimentele efectuate sugerează că prima explicație este mai plauzibilă.

Unitățile de selecție

Selecția naturală nu este doar oportunistă și mioapă, ci, cu doar două excepții pe care le voi discuta imediat, și extrem de individualistă. Selecția nu favorizează specia sau populația, ci organismul individual. Dacă o proprietate care apare dintr-o mutație crește adecvarea relativă a organismului, aceasta se va fixa în populație chiar dacă produce o descreștere a adecvării absolute. Ne putem imagina o populație de pești care este expusă unor prădători în condițiile în care, inițial, înota în bancuri răzlețite. Dacă o mutație îl determină pe peștele la care apare să se deplaseze în centrul bancului, acesta va fi mai puțin vulnerabil la atacurile prădătorului și, prin urmare, va tinde să lase mai mulți urmași. Pe măsură ce comportamentul se răspândește în populație, bancul va deveni din ce în ce mai compact și, astfel, va fi o țintă mai ușoară pentru prădători. În oricare pas al procesului, este mai bine să cauți centrul bancului, și nu marginea acestuia. Însă, în termeni de adecvare absolută, rezultatul este mult mai rău decât situația inițială, iar în termeni de adecvare relativă, nimic nu se schimbă. Similar, instabilitatea selecției sexuale este o explicație plauzibilă pentru coarnele mari și disfuncționale care se regăsesc la anumite specii de cerbi.

O excepție de la individualism este selecția prin înrudire (o formă de „subindividualism"), în care gena, și nu organismul individual, este unitatea de selecție. Alegerea unei unități nu contează când efectul unei gene crește simultan și în aceeași proporție prezența genei în populație și numărul de urmași ai organismului care manifestă acel comportament. Cu așa ceva avem de-a face, de exemplu, în cazul evoluției cu privire la hrănirea mai eficientă. Dar, în anumite cazuri, gena poate beneficia, chiar dacă organismul în care declanșează comportamentul nu beneficiază de pe urma evoluției, și anume când organismul se „sacrifică" pentru binele rudelor apropiate care au șanse foarte mari de a fi purtătoare ale respectivei gene. Când un animal observă un prădător și emite un semnal de alarmă, șansele sale de a supraviețui se diminuează, dar cresc șansele de supraviețuire ale rudelor din preajmă. Din moment ce aceste rude, sau unele dintre ele, sunt purtătoare ale „genei de avertizare", șansele lor mai mari de supraviețuire pot determina gena să se răspândească la nivelul populației dacă reușește să compenseze șansele mici pe care le are animalul care emite semnalul de avertizare. Aceasta nu este singura explicație a acestei forme răspândite de comportament, cu toate că, în anumite cazuri, este cea mai plauzibilă. În alte cazuri, semnalele de alarmă au rolul de a distrage atenția prădătorilor sau de a da un semnal pentru toți membrii

speciei care îl pot auzi (fie ei rude sau nu). În alte cazuri, semnalul poate avea rolul de a-i *înșela* pe alți membri ai speciei, de exemplu, să îi facă să se deplaseze dintr-o zonă unde resursele sunt limitate. Prin urmare, nu este nimic intrinsec altruist sau care să implice sacrificiul de sine cu privire la semnalele de alarmă, cu toate că acestea pot, în anumite circumstanțe, să „imite" altruismul autentic.

O altă excepție este *selecția de grup* (o formă de „supraindividualism"). Să luăm în considerare două populații de pești, una în care mutația prin care se caută centrul bancului a apărut și alta în care nu a apărut. În timp, cea dintâi va lăsa mai puțini urmași decât cea din urmă, iar, în cele din urmă, va fi dislocată. Selecția pare să opereze la nivelul grupului, nu al individului. Însă, dacă cele două populații coexistă, cea de a doua este vulnerabilă la invadatorii din prima specie. În timp ce comportamentul de căutare a centrului este determinat de mutație sau de migrare, rezultatul este același, anume dislocarea acelora care nu se comportă astfel. Asemănător, dacă organismele unei populații au o genă care le împiedică să pască excesiv evitând astfel „tragedia bunurilor publice", acestea pot fi dislocate prin reproducere de organisme mai puțin inhibate cărora le lipsește această genă. Din acest motiv, selecția de grup nu este considerată un mecanism plauzibil pentru a genera cooperare sau autocontrol. Însă, în lumina teoriei pedepsei altruiste pe care am enunțat-o în capitolul 15, putem da un răspuns acestei obiecții. Dacă organismele unei populații au o genă care le face să îi pedepsească pe aceia care nu cooperează, cei din urmă nu vor obține niciun avantaj reproductiv din parazitismul lor. Însă acest mecanism poate doar să asigure cooperarea, acolo unde cei care nu cooperează pot fi identificați destul de sigur. Pare greu de crezut că li se poate aplica peștilor care manifestă comportamentul căutării centrului, dar ar putea foarte bine să se aplice acelor animale care refuză să împartă hrana cu altele.

Selecția de grup și prin înrudire furnizează două mecanisme pentru apariția comportamentului cooperant, primul bazat pe gene împărtășite, iar cel din urmă pe pedeapsa altruistă. Un al treilea mecanism este acela al altruismului reciproc sau „serviciu contra serviciu" în acțiuni repetate, de tipul „îți ofer hrană dacă am un surplus; îmi vei oferi și tu atunci când vei avea mai mult". Reversul medaliei este pedeapsa sau, cel puțin, abținerea de la cooperare când cealaltă parte nu cooperează. Pentru ca acest mecanism să funcționeze, indivizii trebuie să interacționeze destul de mult pentru ca autocontrolul să merite, să își aducă aminte ce au făcut ceilalți mai demult și să îi recunoască atunci când îi întâlnesc din nou.

Acest mecanism pare să dea seamă de cooperare în multe cazuri de interacțiune diadică; în cele ce urmează, voi da două exemple și un contraexemplu. Când două maimuțe capucin au fost puse în situația de a-și împărți hrana cu diferite ocazii, cea de-a doua i-a dat primei maimuțe aceeași cantitate suplimentară pe care a primit-o de la aceasta. Cimpanzeii au o tendință mai mare de a-și împărți hrana cu cei care i-au puricat mai devreme în decursul zilei. Când doi grauri sunt puși într-o situație în care

fiecare primește hrană doar dacă unul dintre ei apasă o manetă, aceștia apasă maneta pe rând. Într-un cuib de lilieci-vampir, femelele regurgitează sânge oferindu-l ca hrană acelora care au eșuat în a-și asigura porția de hrană. Vampirii sunt capabili să se recunoască unii pe alții, existând șanse mai mari ca aceștia să le ofere sânge acelora care au oferit la rândul lor în trecut, însă sunt șanse și mai mari să le ofere sânge rudelor. Într-un alt experiment, două gaițe albastre au fost incluse într-un program care oferă recompense dacă pasărea apasă „maneta egoistă" A, mai degrabă decât dacă apasă „maneta cooperantă" B, indiferent de acțiunile celeilalte păsări și ambele sunt recompensate mai mult dacă ambele apasă B decât în cazul în care niciuna nu apasă această manetă. Aici situația de tip serviciu contra serviciu nu apare: după o anumită cooperare inițială, păsările ajung să apese constant maneta A.

Altruismul reciproc nu este, însă, un mecanism plauzibil pentru generarea cooperării la nivelul grupurilor mari. Singura strategie de pedepsire care ar putea funcționa este „declanșarea răspunsului cumplit", în urma căruia trădarea (comportamentul necooperant) al unui membru al grupului îi determină imediat pe ceilalți să înceteze orice cooperare, atât cu acela care trădează, cât și între ei. Intuitiv, această reacție extremă nu pare plauzibilă și nici nu pare a avea loc la nivel empiric.

◈◈◈

Notă bibliografică

În „Selection by consequences", *Science*, 213 (1981), pp. 501-504, B.F. Skinner a susținut importanța a *trei* moduri în care comportamentul poate fi explicat prin consecințele sale: prin funcționarea selecției naturale la nivelul indivizilor, prin consolidare și (cu toate că nu acesta este termenul folosit) prin selecție de grup. O introducere utilă la teoria consolidării o reprezintă J.E.R. Staddon, *Adaptive Behavior and Learning* (Cambridge University Press, 1983). Un studiu asupra modului în care teoria consolidării poate fi folosită pentru a modela (mai degrabă decât pentru a explica) comportamentul este D. Lee și P. Belflore, „Enhancing classroom performance: A review of reinforcement schedules, *Journal of Behavioral Education*, 7 (1997), pp. 205-217. O expunere clasică a teoriei selecției naturale, remarcabilă pentru insistența asupra caracterului individualist al selecției, este G. Williams, *Adaptation and Natural Selection* (Princeton, NJ, Princeton University Press, 1966). Pentru o discuție a urcării pantei și a „metaforei peisajelor adaptive", vezi capitolul 2.4 din S. Gavrilets, *Fitness Landscapes and the Origin of Species* (Princeton, NJ, Princeton University Press, 2004). O abordare care scoate în evidență gena ca unitate de selecție se află în R. Dawkins, *The Selfish Gene*, ediția a II-a (Oxford University Press, 1990). O introducere excepțională la semnalele emise de animale se găsește în S.A. Searchy și S. Nowicki, *The Evolution*

of Animal Communication (Princeton, NJ, Princeton University Press, 2005). Pentru o discuție a modului în care selecția de grup ar putea avea loc ca urmare a pedepsei altruiste, vezi E. Fehr și U. Fischbacher, „Social norms and human cooperation", *Trends in Cognitive Sciences*, 8 (2004), pp. 185-190. Un studiu esențial al cooperării de tip „serviciu contra serviciu" între animale neînrudite se află în R. Axelrod și W. Hamilton, „The evolution of cooperation", *Science*, 211 (1981), pp. 1390-1396.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 17

SELECȚIA ȘI COMPORTAMENTUL UMAN

Variație și selecție

Cât de mult din comportamentul uman este ales sau „selectat" de vreun alt mecanism decât alegerea rațională? Pentru a răspunde la această întrebare trebuie mai întâi să observăm că orice mecanism de selecție are nevoie de materie primă, inputuri, pe baza cărora funcționează. La fel ca procesul selecției însuși, sursele variației pot fi intenționate sau neintenționate (vezi Figura 17.1).

	Sursa intenționată a variației	Sursa neintenționată a variației
Selecție intenționată	Selecția artificială a plantelor și animalelor în agricultură	Îmbunătățirea graduală a ambarcațiunilor Eugenia Avortul selectiv și pruncuciderea
Selecție neintenționată	Selectarea firmelor prin competiția de piață	Selecția naturală (capitolul 16)

FIGURA 17.1

BCU IASI CENTRAL UNIVERSITY LIBRARY

Variația neintenționată, selecția neintenționată

Selecția naturală a modelat în mod evident structura fizică a ființelor umane, astfel încât să le ofere acestora oportunități pentru acțiune, precum și constrângeri impuse acțiunii. Cei care încearcă să explice comportamentul uman în termeni de selecție naturală fac uneori afirmații mai tari. Aceștia vor să explice modelele comportamentale în sine, nu doar structurile care le fac posibile.

Mecanismul cel mai plauzibil este acela că evoluția a produs *emoții*, cu tendințele lor caracteristice spre acțiune. Din moment ce un mascul nu poate fi niciodată sigur că este tatăl urmașilor săi, în timp ce o femeie nu are vreodată îndoieli în această privință, ne-am aștepta ca selecția naturală să producă o tendință mai mare de a simți gelozie în cazul bărbaților decât al femeilor. Lucrul este confirmat și de statisticile referitoare la sinucideri. Astfel, din 1 060 de sinucideri ale unor persoane căsătorite în Canada între 1974 și 1983, 812 au fost comise de bărbați și 248 de femei, dar, printre cele motivate de gelozie, 195 aparțin bărbaților și doar 19 femeilor. Teoria selecției naturale prezice, de asemenea, că părinții ar fi mult mai dedicați din punct de vedere emoțional copiilor lor biologici decât copiilor vitregi întrucât cei dintâi sunt purtătorii genelor lor. Și această predicție este confirmată de date empirice. Astfel, un copil american care trăia alături de un părinte vitreg în 1976 avea un risc de 100 de ori mai mare de a fi abuzat decât un copil care trăia alături de părinții săi naturali. Selecția naturală poate, de asemenea, să favorizeze lipsa emoției. Pericolele consangvinizării sunt ținute sub control, atât la oameni, cât și la alte specii, prin lipsa unei atracții sexuale între tinerii care cresc împreună, fie că aceștia sunt înrudiți, fie că nu.[1]

Selecția naturală, manifestându-se mai mult la nivelul grupurilor decât la nivelul indivizilor, poate fi responsabilă pentru favorizarea mâniei și indignării față de aceia care încalcă normele cooperării, motivând pedepse care presupun chiar un anumit cost pentru acela care aplică pedeapsa (capitolul 16). O chestiune ceva mai paradoxală este dacă și de ce selecția ar fi trebuit să favorizeze emoția *disprețului*, îndreptată mai degrabă înspre aceia care încalcă norme sociale, decât înspre aceia care încalcă norme morale. Din moment ce multe norme sociale sunt arbitrare și chiar disfuncționale (capitolul 22), este dificil să ne dăm seama cum ar putea fi acestea sprijinite de selecția de grup. Dată fiind tendința a celorlalți de a-i ostraciza pe aceia care încalcă norme sociale, adecvarea reproductivă ar fi mai bine servită de respectarea normelor în cauză dacă costul ostracizării ar fi mai mare decât beneficiul desprins din încălcarea normei. De ce ar apărea o astfel de tendință de la bun început? De ce, de pildă, ar trebui oamenii

[1] Tabuul incestului ar putea, prin urmare, să se adreseze unei chestiuni care apare chiar mai rar decât se crede. Dimpotrivă, Freud a crezut că tabuul incestului a apărut pentru a împiedica o dorință inconștientă de a avea relații sexuale cu rude apropiate.

să condamne adulterul? Normele sociale împotriva adulterului presupun reacții ale unor terți care diferă de reacțiile părților care simt emoția geloziei. Cu toate că A poate beneficia de pe urma faptului că C condamnă avansurile lui B față de partenerul de viață al lui A, acel beneficiu nu creează o presiune de selecție asupra lui C pentru a se comporta în acest mod. În timp ce selecția de grup ar putea favoriza gene care determină „pedeapsirea de către terți" a acelora care calcă strâmb, beneficiul pentru grup al acestui tip de pedeapsă în raport cu adulterul este mult mai puțin evident. Cu toate că tendința potrivit căreia normele cu privire la adulter în raport cu femeile sunt mai puternice decât acelea în raport cu bărbații sugerează o explicație evoluționistă, este dificil de înțeles care este mecanismul subiacent.[1]

Alte teze sunt chiar mai speculative, precum ideea că *autoînșelarea* la oameni a evoluat datorită beneficiilor sale evolutive. Argumentul este următorul. Este de multe ori util să îi înșeli pe alții. Însă autoînșelarea a evoluat pentru a le permite oamenilor să îi înșele pe alții cu mai mult succes. Punctul slab al argumentului este acela că, dacă autoamăgirea determină întreținerea unor credințe false, acestea pot avea consecințe dezastruoase în cazul în care sunt folosite ca premise pentru comportament. Nimeni nu a adus argumente convingătoare că efectul *net* al acestor efecte opuse tinde să fie pozitiv, așa cum ar trebui să fie pentru ca autoînșelarea să sporească adecvarea evolutivă.

Chiar mai speculativă este teza conform căreia *depresia* unipolară a evoluat ca un instrument de negociere, oarecum asemănător unei greve muncitorești. De pildă, o pretinsă funcție a depresiei postnatale este de a-i determina pe ceilalți să își asume un rol în creșterea copilului, la fel cum muncitorii care intră în grevă vor să îi determine pe angajatori să împartă profiturile. Sinuciderile determinate de depresie sunt, conform acestei viziuni, costul efectuării unei amenințări credibile de sinucidere. Acestea sunt, cum ar veni, tentative de sinucidere care au eșuat în eșecul lor. Insomnia este explicată ca o alocare de resurse cognitive pentru rezolvarea crizei pentru care depresia este un răspuns, în timp ce hipersomnia (dormitul mai mult decât este nevoie) este explicată ca un mod de a reduce productivitatea, sporind astfel eficiența de negociere a depresiei. Argumentul, cu toate că este consistent cu anumite fapte cunoscute despre depresie, ignoră o sumă de alte fapte, cum ar fi acela că depresia, la fel ca sinuciderea, se transmite genetic, că indivizii divorțați (care nu au partener de negociere) sunt mai predispuși la depresie decât cei căsătoriți sau care nu au fost niciodată căsătoriți, iar evenimentele stresante ale vieții nu sunt nici necesare, nici suficiente pentru a produce depresie.

---

[1] Probabil, însă, că orice sociolog „respectabil" ar fi în stare să găsească o „explicație" în mai puțin de jumătate de oră.

Explicarea depresiei ca instrument de negociere este alt exemplu de căutare comprehensivă a *sensului* sau *funcției* tuturor comportamentelor fără sens sau disfuncționale. Până la un anumit punct, căutarea sensului este o strategie bună de cercetare; dar, dincolo de acest punct, ea devine din ce în ce mai mult o născocire și, ca în cazul unora dintre exemplele menționate, în cele din urmă devine absurdă. Există atât de multe moduri prin care o trăsătură dăunătoare se poate conserva într-o populație, încât nu putem accepta necondiționat faptul că un comportament care apare frecvent conferă agentului adecvare reproductivă.[1] Selecția naturală a favorizat cu siguranță înclinația de a simți durere fizică și nu avem niciun motiv *a priori* pentru care să credem că nu a favorizat și durerea mentală. Dar, pentru a da seamă de funcția depresiei, nu este suficient să oferim o ipoteză fantezistă care să acopere trăsăturile cunoscute ale bolii. În mod crucial, ipoteza trebuie, să explice și un set cu mult mai mare de fapte decât acelea pentru care am construit-o (capitolul 1) și, preferabil, să dea seamă de „fapte noi", care erau necunoscute până în momentul în care au fost prezise de ipoteză.

Variație intenționată, selecție intenționată

Cu toate că acest dublu mecanism intenționat a fost și va rămâne (cel puțin, putem spera) neimportant în explicarea modelelor comportamentului uman, merită, totuși, să-i luăm în considerare structura. În *Originea speciilor* Darwin scrie că „natura dă variații succesive; omul le adună apoi în anumite direcții care îi sunt utile". Însă nu avem de-a face doar cu ceva de felul „natura propune, omul dispune", din moment ce, după cum tot Darwin observa, comportamentul uman poate modifica inputurile:

> Un grad înalt de variabilitate este, desigur, favorabil întrucât pune la dispoziție în mod liber materialul pe care selecția trebuie să lucreze, nu în sensul că simplele diferențe individuale nu sunt mai mult decât suficiente, cu extrem de mare grijă, pentru a îngădui acumularea unei mase mari de modificări în aproape orice direcție dorită. Dar, întrucât variațiile a căror utilitate sau caracter plăcut pentru om apar doar ocazional, șansele ca ele să apară sunt mult sporite prin menținerea unui număr mare de indivizi și, prin urmare, acest lucru devine important în cel mai înalt grad pentru succes. Pe acest principiu, Marshall a remarcat, cu privire la oile din anumite părți din Yorkshire, că, „întrucât acestea aparțin în general unor oameni săraci și se regăsesc în turme mici, nu vor putea fi niciodată îmbunătățite". Pe de altă parte, agricultorii care cultivă cantități mari de plante din aceeași

[1] Pe lângă comportamentul suboptimal produs de întârzieri sau pleiotropie (capitolul 16), acestea pot fi menținute de o varietate de alte mecanisme genetice legate de faptul că organismele care se reproduc sexual au două variante diferite (alele) ale fiecărei gene.

specie au șanse mai mari decât amatorii să aibă succes în obținerea unor soiuri noi și valoroase.

Astăzi putem adăuga că selecția artificială poate fi îmbunătățită prin provocarea de mutații. În plus, menținerea unor „librării genetice" poate preveni reducerea variației genetice, care, altminteri, ar fi un rezultat inevitabil al selectării unor trăsături particulare.

Cu privire la procesul de selecție în sine, Darwin distinge între *două niveluri de intenționalitate*:

> În prezent, cei mai de frunte crescători de animale încearcă selecția metodică, având în minte un scop clar, de a obține urmași noi sau o subspecie superioară oricăreia existente în țară. Dar, pentru scopul nostru, un tip de Selecție, pe care am putea să o numim Inconștientă și care este consecința încercărilor cuiva de a avea în posesie și a împerechea cele mai bune animale, este mult mai importantă. Astfel, cineva care are intenția de a ține prepelicari va încerca să găsească cele mai bune exemplare canine pe care este în măsură să le găsească, iar apoi va împerechea cei mai buni câini pe care îi are la dispoziție, însă fără să aspire la a schimba permanent rasa.

Variație neintenționată, selecție intenționată

Există multe cazuri în care un organism nou sau o formă nouă apare din întâmplare, iar apoi este acceptată sau respinsă pe baza unei alegeri intenționate. În timp ce selecția naturală tinde să producă un număr egal de organisme mascul și femelă, pruncuciderea prin care se favorizează un anumit sex sau, mai nou, avortul care favorizează un anumit sex pot duce la un dezechilibru serios al populației. În India și China „lipsesc" din acest motiv în jur de 80 de milioane de femei. Politica de eugenie a fost utilizată pe scară largă pentru a împiedica reproducerea bolnavilor mintal și a retardaților. În Germania nazistă, între 300 000 și 400 000 de indivizi au fost sterilizați pe un astfel de temei. Întrucât tehnicile de scanare prenatală s-au îmbunătățit, avortul selectiv poate deveni un factor important în determinarea modului în care este alcătuită populația umană. Dacă la un moment dat va fi posibilă chiar determinarea sexului copilului la momentul concepției, selecția va fi deja înlocuită cu alegerea intenționată.

Variația întâmplătoare combinată cu selecția intenționată poate, de asemenea, să influențeze construcția artefactelor. Când ministrul și sociologul norvegian Eilert Sundt a vizitat Anglia în 1862, intrând în contact cu teoria lui Darwin cu privire la selecția naturală (publicată în 1859), a încercat să aplice o versiune proprie pentru construcția de nave:

> Un constructor de nave poate fi foarte abil, însă nu va reuși să construiască vreodată două vase exact la fel, chiar dacă încearcă să facă exact acest lucru. Variațiile care apar în acest mod pot fi numite *accidentale*. Însă chiar și o foarte mică variație este de obicei observabilă în timpul navigării, iar atunci *nu este accidental* faptul că marinarii ajung să *observe* că vasul a fost îmbunătățit sau a devenit mai potrivit pentru scopurile lor și că ar trebui să recomande ca respectivul vas să fie ales sau imitat. (...) S-ar putea crede că fiecare dintre aceste ambarcațiuni este perfectă în felul său întrucât a ajuns la perfecțiune printr-o dezvoltare unilaterală într-o anumită direcție. Fiecare tip de îmbunătățire a progresat până la punctul în care orice altă dezvoltare ar presupune defecte care ar avea mai multă greutate decât avantajele. (...) Mă gândesc că acest proces are loc în modul următor: după ce ideea unei forme noi și îmbunătățite a apărut prima dată, a urmat *o serie lungă de experimente prudente*, care presupunea fiecare schimbări minore, iar acest lucru a dus la rezultatul fericit că vasul ieșit din șantierul constructorului a ajuns să fie pe placul tuturor.

În acest text, Sundt îmbunătățește teoria lui Darwin dintr-un punct de vedere esențial.[1] În timp ce Darwin își mărturisește ignoranța cu privire la originea variației, Sundt dă peste ideea că sursa variației este *eroarea de replicare*, asemănătoare cu erorile tipografice și (cu ce știm acum a fi) mutațiile ADN-ului. Imperfecțiunile constructorului de nave – inabilitatea sa de a construi copii perfecte – este o condiție a perfecțiunii finale a rezultatului final. Sundt observă cu grijă că rezultatul procesului este un punct de maximum local, din care nu se mai poate face vreo îmbunătățire prin schimbări incrementale. În ultima propoziție citată, acesta sugerează că procesul se poate transforma în selecție artificială, în cazul în care oamenii se angajează în experimente deliberate nelăsând variația să apară ca urmare a șansei. La fel ca Darwin, acesta sugerează că inteligența sau intenționalitatea poate apărea la două niveluri: în primul rând, când oamenii *observă* că un model este mult mai bun decât cel precedent, apoi, când aceștia *înțeleg* că îmbunătățirile ar putea fi accelerate dacă variațiile rezultate ca urmare a șansei ar putea fi înlocuite de experimente sistematice.[2]

---

1 Îmbunătățirea a fost posibilă, firește, nu doar pentru că acesta abordează o altă problemă, din moment ce în 1862 nimeni nu avea armătura conceptuală pentru a putea să-și imagineze că sursa variației în organisme poate sta în greșelile întâmplătoare de replicare. Acest salt a fost posibil doar după ce Mendel a demonstrat natura discretă a unităților moștenite (genele), iar Watson și Crick au demonstrat că replicarea este implicată în procesul de transmitere a caracterelor genetice. Mă întreb ce ar fi răspuns Darwin dacă Sundt l-ar fi întrebat dacă sursa variației biologice nu ar putea fi imperfecțiunea mașinăriei reproductive.

2 În acest mod, s-ar putea preveni situația nefericită care ar putea apărea dacă constructorii ar deveni atât de buni încât să nu facă niciodată vreo greșeală.

Variație intenționată, selecție neintenționată

Funcționarea piețelor în economie are anumite trăsături comune cu selecția naturală. Analogia are două versiuni, una relativ apropiată de selecția naturală, iar alta mai îndepărtată. Acestea împărtășesc premisa că, date fiind multiplele limitări ale raționalității umane, firmele sau managerii sunt ineficienți, în sensul că aceștia nu sunt capabili să calculeze deciziile de producție și de marketing care le vor maximiza profiturile. Cu toate acestea, mecanismul funcționării pieței va elimina firmele ineficiente, astfel încât, în orice moment, se vor observa doar firmele eficiente. Totul se întâmplă „ca și cum" managerii ar fi eficienți.

În prima versiune, care este și cea mai simplă, toate firmele încearcă în permanență să își sporească profiturile prin procese de imitație și inovare. Cu toate că imitația în sine nu generează noi inputuri cu care să poată lucra selecția, imitarea *imperfectă* poate să aibă acest rezultat, după cum am observat deja. Firește, inovația este, prin definiție, o sursă de noi inputuri. Când – prin noroc pur – inovația sau imitația imperfectă îngăduie unei firme să producă la costuri mai mici, ea poate vinde la un preț mai mic decât valoarea produselor, aducând concurența la faliment dacă aceasta nu va reuși să adopte noile metode, mai eficiente. Prin oricare mecanism, fie faliment, fie imitație, aceste tehnici eficiente se vor răspândi în populația de firme. Dacă presupunem că atât imitația, cât și inovația apar predominant în pași mici, iar competiția are loc într-un mediu relativ constant, va rezulta un punct de maximum local al profiturilor de echilibru.

A doua versiune neagă faptul că firmele încearcă mereu să își maximizeze profiturile. Mai degrabă, acestea folosesc *practici de rutină* sau reguli empirice care sunt menținute atâta timp cât profiturile se află la un nivel „satisfăcător". Ca să folosim un neologism, avem de a face cu satisfacere, și nu cu maximizare. Semnificația acestui lucru poate depinde de mulți factori, dar putem presupune, de dragul simplificării, că o firmă ale cărei profituri se află constant sub nivelul satisfăcător fie va da faliment, fie va fi pusă în fața unei amenințări de preluare ostilă. Cea mai simplă rutină este de a face la fel ca înainte, atâta timp cât profiturile se mențin la un nivel satisfăcător. Rutine mai complicate sunt acelea care presupun fixarea unor prețuri printr-o ridicare permanentă a costurilor sau prin investirea unui anumit procent din profituri în producție nouă. Ideea de satisfacere este reflectată în maxime precum „Nu schimba niciodată echipa învingătoare" sau „Dacă nu e stricat, nu ai ce să repari". Dintr-o anumită perspectivă, satisfacerea ar putea fi chiar optimă. Pentru a relua o expresie citată mai devreme, „Cel mai mare profit pe care îl aduce monopolul este o viață liniștită".

Să presupunem acum că profiturile scad sub nivelul satisfăcător. O firmă care face exact același lucru de ani și ani poate fi victima unui echivalent organizațional al ruginirii sau sclerozării. Șocurile externe, precum o creștere a prețului petrolului sau o modificare importantă a ratei de schimb, pot crește costurile sau reduce veniturile.

Cererea consumatorilor se poate schimba, concurența poate dezvolta metode mai competitive pentru a produce bunuri noi, muncitorii pot impune o grevă costisitoare în firmă. Indiferent de cauză (care ar putea rămâne necunoscută chiar firmei), profiturile nesatisfăcătoare vor determina căutarea unor practici de rutină noi printr-o combinație de inovație și imitație. Sunt șanse mari ca fiecare procedură să aibă un caracter predominant local, în sensul de a fi limitată la alternativele apropiate de practicile de rutină existente. Schimbările majore de orice tip pot fi prea costisitoare pentru o firmă care are probleme financiare (capitolul 9), iar inovațiile neincrementale presupun un efort de proiectare mult mai mare.

Procesul imitării este, în mod evident, părtinitor față de comportamentul competitorilor care au succes. Dacă că inovația este întâmplătoare sau direcționată, acest lucru depinde de cauzele percepute ale crizei care a declanșat inovația. Dacă scăderea profiturilor sub nivelul acceptabil este rezultatul unei creșteri a prețului la țiței, firma poate fi părtinitoare în căutarea unor soluții în direcția unor metode care să economisească petrol.[1] În cazul în care criza este rezultatul unei modificări a ratei de schimb dintre dolar și euro, sunt șanse mult mai mari ca firma să caute la întâmplare. În orice caz, însă, există o componentă intenționată puternică în comportamentul firmei. Decizia de a schimba practicile de rutină curente este intenționată, ca și decizia despre cât de mult trebuie investit pentru inovație sau în imitație. Alegerea modelelor de imitat este deliberată și, după cum am observat deja, firma poate favoriza o anumită direcție în căutarea unei soluții.

Practicile de rutină noi care rezultă în urma acestui proces sunt expuse apoi forțelor oarbe ale competiției de piață. Dacă acestea îi permit să obțină un nivel satisfăcător al profitului, firma va stopa căutările până se va ciocni de o nouă criză. În caz contrar, firma poate încerca din nou sau va trebui să se declare falită. Mai devreme sau mai târziu, firmele care nu reușesc să ajungă la un nivel satisfăcător al profitului vor fi eliminate. În sine, acest proces nu tinde să producă firme care caută să își maximizeze profiturile. Pentru a vedea cum s-ar putea întâmpla așa ceva, trebuie să descriem mai clar *competiția*. Dacă presupunem că una dintre practicile de rutină ale firmei este aceea de a investi un procent fix al profiturilor în producția de bunuri noi, firmele care, prin simplu noroc, dau peste o practică de rutină mai bună decât concurența se vor extinde astfel încât, în timp, practicile lor vor fi mai bine reprezentate în populația de firme.[2]

---

[1] Dacă nu sunt așteptate creșteri ulterioare, raționalitatea nu impune firmei să caute *inovații pentru a economisi* petrol din moment ce nimeni nu știe cum arată setul de inovații fezabile (vezi, de asemenea, și capitolul 14). Însă creșterea prețului la țiței va tinde să facă aceste inovații mai semnificative.

[2] În această versiune, acestea nu vor încerca să își scoată rivalii de pe piață, de exemplu, folosindu-și profitul pentru a vinde sub prețul pieței până la înfrângerea concurenței întrucât nu sunt preocupate de profituri mai mult decât satisfăcătoare.

Modele de selecție în științele sociale

Utilitatea acestor modele depinde de o simplă întrebare empirică: care este rata la care sunt eliminate firmele ineficiente, în comparație cu rata de schimbare a mediului? Am pus aceeași întrebare în capitolul anterior, dar cu referire la selecția naturală, și am argumentat în mod indirect că adaptarea foarte rafinată a organismelor la mediu sugerează că mediul s-a schimbat relativ lent. În cazul mediului economic, putem face o evaluare mai directă. În lumea modernă, firmele sunt expuse unor procese de schimbare care decurg mai rapid ca niciodată. Dacă ar fi reduse la sincronizarea incrementală cu mediul, firmele ar fi incapabile cronic să supraviețuiască. Cele mai mari șanse de succes le au firmele care sunt capabile să *anticipeze* schimbarea, ochind în fața țintei. Și această strategie va eșua în mare parte a timpului, însă nu chiar tot timpul. În plus, datorită influenței lor în zona politicului, marile corporații sunt capabile să *modeleze* mediul în care funcționează. Într-o epocă timpurie a capitalismului sălbatic practicat de firmele mici, mecanismele de selecție de tipul acelora pe care le-am descris se poate să fi fost importante, dar nu neapărat – nu avem de unde ști. Astăzi, însă, este puțin probabil că acestea ar putea explica ceva din ce vedem că se întâmplă.

Există și o chestiune mai generală în joc. Când sunt criticați pentru lipsa de realism a presupozițiilor lor, teoreticienii alegerii raționale susțin de obicei că încearcă să explice comportamentul pornind de la ipoteza că oamenii acționează „ca și cum" și-ar maximiza utilitatea (ori profitul sau orice alt scop). De multe ori, ei adaugă că această presupoziție nu poate fi justificată de vreun soi de mecanism de selecție. În sfera economică se presupune că, de fapt, competiția de piață ar trebui să fie responsabilă pentru așa ceva. Am argumentat deja că nu este posibil.[1] Poate că modul cel mai general de a-mi enunța obiecția este că, deși am putea arăta că piața, prin competiție, îmbunătățește eficiența eliminând firmele ineficiente, e o cale lungă de la „îmbunătățirea eficienței" până la ultrasofisticata noțiune a maximizării de tip „ca și cum" atribuită firmelor de modelele economice.

În sectorul politic, competiția electorală se presupune că garantează faptul că singurii politicieni pe care îi observăm sunt cei aleși sau realeși; prin urmare, putem presupune că toți politicienii se comportă „ca și cum" ar fi preocupați doar de perspectivele pe care le au de a câștiga alegerile. Saltul de la o anumită preocupare pentru alegeri la o preocupare *exclusivă* pentru victoria în alegeri nu este, însă, justificată. Dacă analizăm politicul fără prejudecăți metodologice, vedem că există trei tipuri de

---

[1] Pe lângă argumentul general pe care l-am menționat, gestionarea echipelor sportive oferă o posibilă obiecție la ideea că maximizarea profiturilor este generată de selecție. Dacă echipele de fotbal sau baseball și-ar folosi profitul, în vederea maximizării sale, pentru a cumpăra cei mai buni jucători din divizie, superioritatea lor ar deveni atât de covârșitoare, încât jocurile și-ar pierde mult din incertitudine și, prin urmare, din spectaculozitate, ceea ce ar determina o diminuare a capacității lor de a genera profit.

actori politici: oportuniștii (cărora nu le pasă decât să fie aleși), reformatorii (cărora le pasă ca strategiile lor să fie implementate) și activiștii sau militanții (cărora le pasă doar de a „transmite un mesaj").[1] Abordarea politicului care se bazează pe interacțiunea dintre aceste trei grupuri la nivelul fiecărui partid – și între partide diferite – este evident mai realistă decât modelul politic de tip „stand de înghețată" (capitolul 19), conform căruia partidele care urmăresc maximizarea voturilor ar converge toate către centru. Pentru un contraargument frapant al tezei conform căreia politicienii sunt motivați doar de preocuparea de a fi realeși, putem lua în considerare suita de politicieni francezi inaugurată de Jean Jaurès și care îi include pe Léon Blum, Pierre Mendès-France și Michel Rocard, toți aceștia motivați în mod transparent de o dorință de a promova valorile imparțiale ale dreptății sociale și ale eficienței economice. Trebuie spus însă că, în cazul lui Rocard, aversiunea acestuia față de politica electorală i-a afectat eficacitatea politică.

În afara arenelor competiționale, raționalitatea lui „ca și cum" are chiar mai puține temeiuri. Alegerile consumatorilor, comportamentul electoral, frecvența mersului la biserică, alegerea unei cariere și cele mai multe comportamente la care ne putem gândi nu sunt supuse mecanismelor de selecție care mimează raționalitatea. Sunt, firește, supuse și ele unor *constrângeri* care pot reduce importanța alegerii în general și a alegerii raționale în special (capitolul 9). Constrângerile funcționează *ante factum* pentru a face anumite alegeri nefezabile. Selecția operează *post factum* pentru a-i elimina pe aceia care au făcut anumite alegeri. Cu toate că ambele mecanisme contribuie la explicarea comportamentului, ele nu pot da seamă de acesta nici împreună, nici separat. Alegerea rămâne conceptul fundamental în științele sociale.

◈◈◈

Notă bibliografică

Datele statistice cu privire la crimă și abuzarea copiilor provin din M. Daly și M. Wilson, *Homicide* (New York, Aldine de Gruyer, 1988). Pentru obiecții la aceste explicații, vezi capitolul 7 din D. Buller, *Adapting Minds* (Cambridge, MA, MIT

[1] Putem face anumite distincții cu privire la cele trei grupuri în modul următor. Oportuniștii preferă să propună politica A în locul politicii B când probabilitatea de a câștiga cu A este mai mare decât probabilitatea de a câștiga cu B, dat fiind faptul că partida adversă propune o anumită politică fixă C. Militanții preferă să propună A lui B când membrul mediu de partid ar obține o mai mare utilitate la A decât la B (indiferent de C). Reformiștii preferă să propună A lui B, dat fiind faptul că opoziția propune C, când utilitatea așteptată a membrului de partid mediu este mai mare la A decât la B. Astfel, oportuniștii sunt preocupați doar de probabilități, activiștii doar de utilități, iar reformatorii de ambele.

Press, 2005). Pentru două fațete ale argumentului autoînșelării, vezi R. Trivers, *Social Evolution* (Menlo Park, CA, Benjamin-Cummings, 1985) (care favorizează o explicație evoluționistă), și V.S. Ramachandran și S. Blakeslee, *Phantoms in the Brain* (New York, Quill, 1998) (pentru argumentul opus). Pentru două fațete ale naturii adaptive a depresiei, vezi E.H. Haggen, „The bargaining model of depression", în P. Hammerstein (ed.), *Genetic and Cultural Evolution of Cooperation* (Cambridge, MA, MIT Press, 2003) (care favorizează o explicație evoluționistă), și P. Kramer, *Against Depression* (New York, Viking, 2005) (pentru argumentul opus). Analiza pieței în termenii selecției naturale își are originea în A. Alchian, „Uncertainty, evolution, and economic theory", *Journal of Political Economy*, 58 (1950), pp. 211-221. Versiunea sa cea mai sofisticată (care nu sprijină maximizarea de tip „ca și cum") se regăsește în R. Nelson și S. Winter, *An Evolutionary Theory of Economic Change* (Cambridge, MA, Harvard University Press, 1982). Teoria cu privire la satisfacere (*satisficing*) îi aparține lui H. Simon, „A behavioral theory of rational choice", *Quarterly Journal of Economics*, 69 (1954), pp. 99-118. Gestionarea echipelor sportive este subiect al cărții lui D. Berri, M. Schmidt și S. Brook, *The wages of Wins* (Stanford, CA, Stanford University Press, 2006). Distincția dintre oportuniști, reformatori și activiști îi aparține lui J. Roemer, *Political Competition* (Cambridge, MA, Harvard University Press, 2001). Pentru argumente de tip „ca și cum" se poate consulta și nota bibliografică a capitolului 1.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

V
INTERACȚIUNEA

Interacțiunea socială poate lua forme multiple. (1) Rezultatul, pentru fiecare agent, depinde de rezultatele celorlalți. Această interdependență a rezultatelor poate apărea dacă bunăstarea fizică sau psihică a celorlalți îmi afectează propria bunăstare psihică (capitolul 5). (2) Rezultatul pentru fiecare poate depinde de acțiunile tuturor. Această interdependență reflectă cauzalitatea socială generală (capitolul 18), ilustrată în fenomene precum încălzirea globală (aceea pentru care omul este responsabil). (3) Acțiunea fiecăruia depinde de acțiunea (anticipată) a tuturor. Această interdependență este subiectul specific al teoriei jocurilor (capitolele 19 și 20), care integrează în cadrul său conceptual și punctele (1) și (2). (4) Credințele fiecăruia depind de acțiunile tuturor. Această interdependență poate apărea printr-o varietate de mecanisme, cum ar fi „ignoranța pluralistă" sau „cascadele informaționale" (capitolul 23). (5) Preferințele fiecăruia depind de acțiunile tuturor. Această interdependență este, poate, aspectul cel mai puțin înțeles al interacțiunii sociale. Cu toate că abordez anumite aspecte ale problemei în mai multe locuri, în special în capitolul 22, nu propun vreo dare de seamă comprehensivă asupra chestiunii.

Aceste interdependențe pot să apară prin acțiunea descentralizată a indivizilor care nu se află în vreo relație organizată unul cu celălalt (capitolul 24). Însă cea mai mare parte a vieții sociale are, totuși, o anumită structură. Multe rezultate apar prin proceduri de decizie colectivă – susținerea unei poziții, votul și negocierea – prin care grupuri de indivizi ajung să ia decizii care îi obligă unii față de alții (capitolul 25). În cele din urmă, organizațiile funcționează prin reguli care sunt concepute astfel încât să calibreze stimulentele indivizilor cu scopurile întregului sistem (capitolul 26).

BCU IAȘI UNIVERSITY LIBRARY

Capitolul 18

CONSECINȚE NEINTENȚIONATE

Consecințe neintenționate ale comportamentului individual

Lucrurile nu ne ies întotdeauna așa cum intenționăm. Multe evenimente se produc neintenționat. Uneori, cauzele sunt banale, ca în cazul în care apăsăm pe accelerație în loc să apăsăm pe frână sau apăsăm butonul „delete" din greșeală. Cu toate că nu poate exista o „teorie generală a consecințelor neintenționate", putem începe prin a întocmi un catalog. Voi lua în considerare cazuri în care consecințele sunt nu doar neintenționate, ci și neprevăzute. „Efectele secundare previzibile ale unei acțiuni" nu sunt intenționate în sine, mai ales dacă sunt negative, dar nu le voi trata drept „consecințe neintenționate ale acțiunii".

Consecințele neintenționate pot apărea atât din comportamentului individual, cât și din interacțiunea socială. Începând cu primul, putem folosi o simplă extindere a cadrului dorință-oportunitate pe care l-am descris amănunțit în capitolul 9 (vezi Figura 18.1).

Cu toate că acțiunile sunt modelate de dorințe (nu de preferințe), acestea pot, la rândul lor, să modeleze dorințele. Astfel, pe lângă rezultatele intenționate ale unei acțiuni, există uneori și unul neintenționat: o schimbare a dorinței. Dependența este un bun exemplu. Sub influența unor droguri care dau dependență, oamenii încep să deprecieze mai puternic viitorul, slăbind astfel efectul descurajant al răului pe care îl aduce dependența pe termen lung. Dacă acest efect ar fi fost anticipat, l-ar fi putut

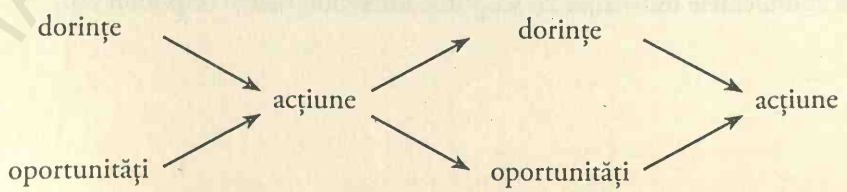

FIGURA 18.1

BCU IAȘI CENTRALĂ UNIVERSITY LIBRARY

face pe agent să nu o ia pe calea dependenței, însă, în mod obișnuit, acest lucru nu se întâmplă. Fenomene asemănătoare se observă și în situații mai normale. Merg la o petrecere intenționând să beau doar două pahare pentru a putea să mă întorc acasă cu mașina, dar, după al doilea pahar, decizia mea se dizolvă în aburii alcoolului și mai beau un pahar. Dacă aș fi știut, aș fi putut să beau doar un pahar.

Dorințele pot fi și ele afectate de o preferință inconștientă pentru noutate sau pentru schimbare („Iarba e mai verde de cealaltă parte a gardului"). În povestea *Tata face întotdeauna bine ceea ce face*, a lui H.C. Andersen, fermierul merge la târg dis-de-dimineață pentru a-și vinde calul sau a face troc cu acesta. Se întâlnește prima dată cu un om care are o vacă, de care îi place atât de mult, încât o schimbă cu calul. În tranzacții succesive, vaca este schimbată cu o oaie, oaia cu o gâscă, gâsca cu o găină, iar, în cele din urmă, găina este schimbată cu un sac de mere stricate. Drumul fermierului spre ruină este pavat cu îmbunătățiri treptate. Fermierul crede de fiecare dată că are ceva de câștigat de pe urma schimbului, dar rezultatul net al tuturor tranzacțiilor este dezastruos.[1]

„Efectul împroprietăririi", o implicație a aversiunii față de pierdere (capitolul 12), ilustrează, de asemenea, schimbările de preferințe determinate de alegere, dar neintenționate.[2] Multe bunuri dobândesc o valoare subiectivă mai mare pentru proprietarul lor decât aveau înainte de a le cumpăra, lucru demonstrat de faptul că prețul minim de vânzare depășește în mod normal prețul maxim de cumpărare cu un factor cuprins între 2 și 4. Din moment ce marea parte a bunurilor este evaluată ca o pierdere când este cedată și ca un câștig când este dobândită și din moment ce pierderile au un impact mai mare decât câștigurile de dimensiuni comparabile, aversiunea față de pierdere prezice acest lucru. Mai mult, experimentele arată că potențialii cumpărători subestimează prețul minim de revânzare pe care l-ar accepta, ceea ce indică faptul că schimbarea

---

[1] Formalizând, ne putem imagina o persoană care, cu regularitate (cu toate că nu conștient), își ajustează dorințele sale astfel încât preferă mai mult bunul din care are în mod curent mai puțin. Să presupunem că este expus următorului șir de pachete de două bunuri (1/2, 3/2), (3/4, 1/2), (1/4, 3/4), (3/8, 1/4)... Apoi, dacă, la un anumit moment, acesta consumă pachetul n din șir și, pentru perioada următoare, i se oferă o alegere între pachetul n și pachetul $n + 1$, acesta va alege întotdeauna cel din urmă pachet din moment ce oferă mai mult din bunul pe care în mod curent îl are într-o cantitate mai mică. Dar, întrucât șirul converge către zero, aceste îmbunătățiri locale vor aduce, per total, ruina. Efectul este asemănător cu acela în care un agent poate fi adus în sapă de lemn prin preferințe ciclice (capitolul 11), dar mecanismul este diferit.

[2] Din întâmplare, termenul „efect al împroprietăririi" a ajuns să fie folosit atât pentru tendința de a supraevalua lucrurile posedate, cât și pentru utilitatea pe care o persoană o poate obține în prezent din utilitatea din trecut (capitolul 2). Cele două semnificații nu au nicio legătură între ele.

de preferințe este, într-adevăr, neprevăzută.[1] Un alt mecanism care ar putea produce acest „efect de întărire", adică tendința de a vedea alegerile personale într-o lumină favorabilă odată ce au fost făcute, este oferit de teoria disonanței cognitive (capitolul 1).

Un exemplu al modului în care acțiunea poate modela *oportunitățile* în moduri neintenționate și neprevăzute este oferit de cazul unui bătăuș în măsură să își impună punctul de vedere în tranzacțiile cu ceilalți pentru că aceștia preferă în general să cedeze, decât să îl înfrunte. Una dintre consecințele neintenționate ale comportamentului său poate fi aceea că este ocolit de ceilalți, astfel încât șansele să aibă tranzacții cu respectivele persoane vor fi mai mici. Bătăușului îi iese de minune fiecare întâlnire, însă acestea sunt rare. Această consecință poate fi nu doar neintenționată și neprevăzută, ci și neobservată. Din câte poate bătăușul să își dea seama, practicile sale funcționează.[2] Dacă reușește să observe efectele negative ale comportamentului său, acesta poate să continue pe aceeași linie dacă efectele pozitive au mai multă greutate decât cele negative. În acest caz, consecințele negative vor fi prevăzute, dar nu intenționate în sine.

De multe ori, alegerea unei opțiuni astăzi va îndepărta anumite opțiuni din setul fezabil în viitor. Acest efect poate fi prevăzut: constrângerile asupra bugetului meu îmi permit să-mi cumpăr o mașină, dar nu două. Uneori, însă, agentul poate să nu știe că alegerea sa are consecințe ireversibile. Un țăran poate avea o bucată de teren pe care se găsesc niște copaci și câteva parcele cultivabile. Pentru a obține mai mult teren pentru cultivat și lemne de foc, taie copacii. Defrișarea cauzează eroziune, prin urmare, în timp, va ajunge să aibă chiar mai puțin teren cultivabil decât la început. Într-o serie de cazuri, pe care le voi discuta în cele ce urmează, eroziunea poate fi rezultatul unui comportament *colectiv* în cazul în care, de exemplu, se produce pe terenul fermierului dacă și numai dacă atât fermierul, cât și cei doi vecini efectuează defrișări. Însă este posibil și destul de obișnuit ca un singur individ să își submineze fără să știe oportunitățile viitoare pentru acțiune. Responsabil pentru aceasta este *deficitul cognitiv*: agentul nu poate prezice consecințele viitoare ale comportamentului prezent. În alte cazuri, se poate produce printr-un *deficit motivațional*: agentul nu acordă prea mare importanță consecințelor viitoare (cunoscute și sigure) în raport cu câștigurile imediate (capitolul 6).

---

[1] Într-adevăr, este dificil să ne dăm seama cum ar putea fi prevăzut. Persoana ar trebui să folosească două puncte de referință diferite simultan, starea de dinainte de cumpărare în funcție de care bunul este un câștig și starea de după cumpărare în funcție de care este o pierdere.

[2] Această perspectivă limitată este împărtășită de unii cercetători din domeniul științelor sociale care susțin că emoții precum mânia pot fi „raționale" sau, cel puțin, adaptive pentru că le permit agenților să își impună voia în tranzacțiile cu ceilalți.

Externalitățile

Mă voi ocupa în cele ce urmează de consecințele neintenționate ale *interacțiunii*, o temă care, la origine, a fost o idee esențială a științelor sociale. Ca să citățm fraza memorabilă a lui Adam Ferguson, istoria este „rezultatul acțiunii umane, dar nu urmează punerea în practică a unui plan uman". Contemporanul său, Adam Smith, se referă la o „mână invizibilă" care modelează relațiile umane. O jumătate de secol mai târziu, Hegel avea să invoce „viclenia rațiunii" pentru a explica progresul libertății în istorie. Cam în același timp, Tocqueville enunța o teză similară, conform căreia, în progresul democrației, „toată lumea joacă un rol: cei care se străduiesc să asigure succesul democrației, precum și cei care nici nu au visat vreodată de a-i fi supuși; cei care s-au luptat pentru, ea precum și cei care i-au declarat război". Câțiva ani mai târziu, Marx avea să se refere la „alienarea" oamenilor de propria lor acțiune, susținând că „această fixare a activității sociale, această consolidare a ceea ce producem noi înșine ca o putere materială deasupra noastră, care crește și ne scapă de sub control, ne dejoacă așteptările, face pulbere toate planurile noastre, este unul dintre factorii esențiali în dezvoltarea istorică de până acum".

Dintre toți acești autori, doar Adam Smith și Marx furnizează mecanisme specifice prin care se produc consecințe neintenționate. În limbaj modern, ei scot în evidență modul în care externalitățile comportamentului se pot cumula pentru a produce rezultate care nu sunt nici intenționate, nici prevăzute de către agenți. În formă stilizată, ne putem imagina că fiecare agent dintr-un mare grup de agenți identici face o anumită acțiune pentru a-și promova propriul interes. Ca rezultat secundar al acelei acțiuni, agentul impune un anumit cost mic sau conferă un anumit beneficiu mic (externalitate negativă sau pozitivă) asupra tuturor celorlalți agenți (inclusiv asupra sa). Fiecare agent este ținta multor astfel de acțiuni. Adunând efectele și însumând apoi beneficiile private ale agentului cauzate de acțiunea sa, obținem rezultatul final pe care îl generează agenții prin acțiunile lor. Din moment ce presupunem că aceștia sunt identici, stările lor inițiale, stările pe care aceștia intenționează să le realizeze și stările pe care le realizează de fapt în mod colectiv pot fi reprezentate printr-un singur număr x, y, și, respectiv, z.[1]

Să presupunem mai întâi că $z > y > x$, ceea ce reprezintă o externalitate pozitivă. Acesta era interesul central al lui Adam Smith: când un agent își direcționează

[1] Mulți economiști nu ar include în rândul externalităților unele dintre fenomenele pe care le menționez eu aici. Ar include poluarea, dar nu efectele generate de piață, precum șomajul keynesian. Însă, având în vedere scopul lucrării de față, ceea ce contează este elementul lor comun: în căutarea unui beneficiu pentru sine, fiecare individ impune un mic cost sau beneficiu asupra tuturor celorlalți *și asupra sa*. O firmă care concediază angajați sau taie lefurile va determina o mică reducere a cererii pentru propriile produse.

„industria în așa fel încât produsul său să fie de cea mai mare valoare, acesta are în vedere propriul său câștig și este condus în acest demers, ca și în alte cazuri, de o mână invizibilă, care promovează un scop care nu este parte a intenției sale. Și nici pentru societate nu este întotdeauna mai rău dacă interesul său nu include așa ceva. Urmărindu-și propriile interese, acesta promovează adesea mai eficient interesul societății decât atunci când vrea să îl promoveze în mod intenționat." În competiția de piață scopul fiecărei firme este acela de a face profit producând mai ieftin decât rivalii, dar, procedând în acest mod, urmăresc avantajul consumatorilor. Și clienții ar putea, în calitatea lor de muncitori sau manageri, să fie în situația asemănătoare, de a aduce beneficii altora prin eforturile lor competitive. Rezultatul a fost o creștere spectaculoasă, vreme de secole. Poate că efectul a fost sau nu prevăzut, dar în niciun caz nu a fost „parte a" intenției cuiva.

Să presupunem acum că $y > z > x$, caz în care avem de a face cu o externalitate negativă slabă. Agenților le este mai bine în urma eforturilor depuse, însă, din cauza costurilor pe care și le impun unii altora, nu chiar atât de bine pe cât s-ar fi așteptat. Oamenii care fac naveta la serviciu cu mașina se pot afla într-o situație mai bună decât aceia care folosesc transportul în comun în cazul în care acesta din urmă este slab dezvoltat, dar blocajele în trafic și poluarea îi împiedică să se bucure de situație pe cât ar fi fost de așteptat. Dacă externalitatea este produsă de blocaje, aceștia nu pot să nu o remarce. Dacă, însă, este produsă de poluare, poate să dureze o vreme până să înțeleagă că își fac rău unii altora și că nu sunt victime (doar) ale poluării industriale.

Să presupunem, în cele din urmă, că avem de a face cu o externalitate negativă puternică, caz în care $y > x > z$. Agenții se află cu toții într-o situație mai rea în urma încercării fiecăruia de a ajunge într-o situație mai bună. Aceasta este una dintre acuzațiile principale aduse de Marx la adresa economiei capitaliste descentralizate. Teoria sa principală cu privire la crizele capitaliste, „teoria ratei descrescătoare a profitului", are o astfel de structură generală. Pentru a menține sau pentru a crește profiturile, susține acesta, fiecare capitalist este încurajat să înlocuiască muncitorii cu mașinile. Însă când toți capitaliștii fac acest lucru simultan, își taie împreună craca pe care stau din moment ce sursa profitului, în cele din urmă, este plusvaloarea generată prin muncă. Argumentul este seducător, dar, la o analiză mai atentă, se dovedește greșit din mai multe puncte de vedere. Mai interesantă este altă observație, pe care Marx o face în treacăt și care va deveni ulterior una dintre pietrele de temelie ale teoriei șomajului propusă de John Maynard Keynes. Fiecare capitalist, observă Marx, are o relație ambiguă cu muncitorii. Pe de o parte, acesta vrea ca muncitorii pe care *el* îi angajează să aibă salarii mici, întrucât acest lucru presupune profituri mari. Pe de altă parte, capitalistul vrea ca toți *ceilalți* muncitori să aibă salarii mari, întrucât astfel își asigură o cerere mare pentru produsele sale. Cu toate că, pentru un anumit capitalist, este posibil să

își satisfacă aceste două dorințe simultan, din punct de vedere logic este imposibil ca acest lucru să se întâmple pentru toți capitaliștii. Aceasta este „contradicția capitalismului", pe care Keynes o prezintă în modul următor. Într-o situație în care profiturile scad, fiecare capitalist reacționează prin concedierea angajaților, economisind astfel cheltuielile salariale. Dar, întrucât cererea de muncitori este ceea ce susține firmele, direct sau indirect, efectul produs de reducerile de personal efectuate simultan de către toți capitaliștii ar duce la o mai mare scădere a profiturilor, care ar cauza mai multe disponibilizări sau chiar falimente.

Există multe cazuri care se înscriu în acest tipar general. Pescuitul în exces, defrișarea și pășunatul intensiv („tragedia bunurilor publice") pot fi raționale din punct de vedere individual, dar, din punct de vedere colectiv, sunt suboptime sau chiar dezastruoase. Dacă fiecare familie dintr-o țară în curs de dezvoltare concepe mulți copii pentru a se asigura împotriva sărăciei la bătrânețe, suprapopularea va genera și mai multă sărăcie. Într-o criză a apei, fiecare individ care folosește apa pentru scopuri neesențiale cauzează o ușoară creștere a probabilității ca autoritățile să taie alimentarea cu apă pentru câteva ore în fiecare zi, afectând astfel și scopurile esențiale. Aceste consecințe pot fi prevăzute sau nu. O trăsătură crucială a acestei categorii de consecințe neintenționate este aceea că, și atunci când sunt prevăzute, comportamentul rămâne același. După cum voi explica în capitolul următor, aceasta este *strategia dominantă*: este rațional să alegi, indiferent de ce fac ceilalți.

Internalitățile

Un argument parțial asemănător se aplică și pentru ceea ce putem numi „internalități", definite ca beneficiile sau daunele pe care alegerea unei persoane la un anumit moment le poate conferi bunăstării obținute din alegeri ulterioare. Pentru a folosi o metaforă, putem spune că internalitățile sunt externalități pe care o persoană le impune „sinelui său viitor".

În discuția cu privire la custodia copiilor pe care am rezumat-o în Figura 11.3, am susținut că timpul petrecut alături de copil produce o internalitate pozitivă pentru părinte. Dependența ne dă un alt exemplu important de internalități negative. Cu cât o persoană a consumat mai mult în trecut dintr-o substanță care dă dependență, cu atât va obține mai puțină plăcere în prezent. Acest efect de „toleranță" poate să se producă și în cazul bunurilor care nu dau dependență. Chiar dacă îți place la nebunie un anumit sortiment de înghețată, te vei sătura de ea dacă vei mânca cinci porții pe zi. În cazul dependenței, însă, consumul trecut are și alt efect. În timp ce produce o plăcere mai mică decât dacă agentul nu ar fi consumat nimic în trecut, crește diferența de bunăstare dintre faptul de a consuma în prezent și a nu consuma în prezent („sevraj"). Vezi Figura 18.2 pentru schema acestei situații.

Astfel, indiferent dacă s-a abținut sau a consumat în trecut, agentul se află într-o stare mai bună consumând în prezent decât abținându-se în prezent. Consumul este strategia dominantă. În același timp, consumul repetat duce la o stare mai rea în toate momentele (cu excepția unor foarte puține momente chiar la început) decât abținerea repetată, la fel cum strategia dominantă a tuturor de a avea mai mulți copii îi duce pe toți într-o stare mai proastă. Există, desigur, diferențe evidente între externalități și internalități. Una dintre ele este asimetria temporală: în timp ce toți indivizii își pot face rău unii altora, sinele viitor nu-i poate face rău sinelui prezent. Altă diferență este că sinele viitor nu este decât o porțiune temporală *a unuia și aceluiași* decident, pe când indivizii diferiți nu sunt părți spațial distincte ale unui superorganism. Imediat ce persoana *însăși* înțelege că alegerile sale prezente au un efect negativ asupra bunăstării pe care o poate obține din alegeri ulterioare, aceasta capătă un stimulent pentru modificarea comportamentului. Dacă stimulentul este suficient de puternic, acest lucru depinde de severitatea simptomelor de sevraj și de măsura în care agentul depreciază bunăstarea viitoare. Unii agenți care nu ar fi făcut niciodată primul pas dacă ar fi știut care vor fi consecințele pot alege să nu se lase odată ce sunt prinși.

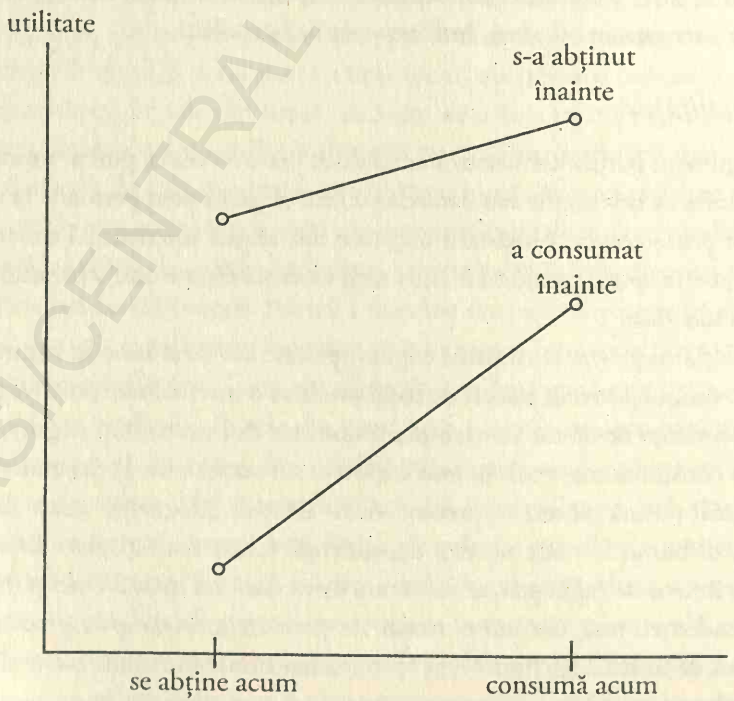

FIGURA 18.2

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Sindromul fratelui mai mic

Consecințele neintenționate ale acțiunii sociale pot fi produse și de ceea ce voi numi *mecanismul fratelui mai mic*. Înainte de a explica această expresie, o voi ilustra cu un exemplu faimos din teoria economică, „pânza de păianjen", numit și *hog cycle* pentru că a fost propusă ca explicație a fluctuațiilor ciclice în producția de porci. Aplicația este mult mai largă, însă. Fluctuațiile în industria constructoare de nave au urmat adesea o traiectorie asemănătoare, cu o piață bună pentru vânzători, urmată de supraînvestire și, apoi, de excedent. Studenții care își aleg cariera pe baza cererii curente de absolvenți pot distruge colectiv însuși temeiul deciziei lor.

Crescătorii de porcine trebuie să se decidă cu un an înainte cât de mult vor să scoată pe piață în anul care vine, o decizie determinată de prețul pe care se așteaptă să îl primească pe cap de animal și de costurile de producție. O creștere a prețului estimat îi va determina pe fermieri să producă mai mult, lucru reflectat în curba ofertei cu pantă ascendentă din Figura 18.3.

Să presupunem că în anul 1 prețul pentru porcine este $p1$. Așteptându-se ca prețurile să rămână la același nivel în anul 2, fermierii împing pe piață $q2$ anul viitor. Însă, la acest volum, prețul de clearing al pieței este $p2$, și nu $p1$. Așteptându-se ca prețurile să rămână la același nivel în anul 3, fermierii vor produce volumul $q3$ pentru

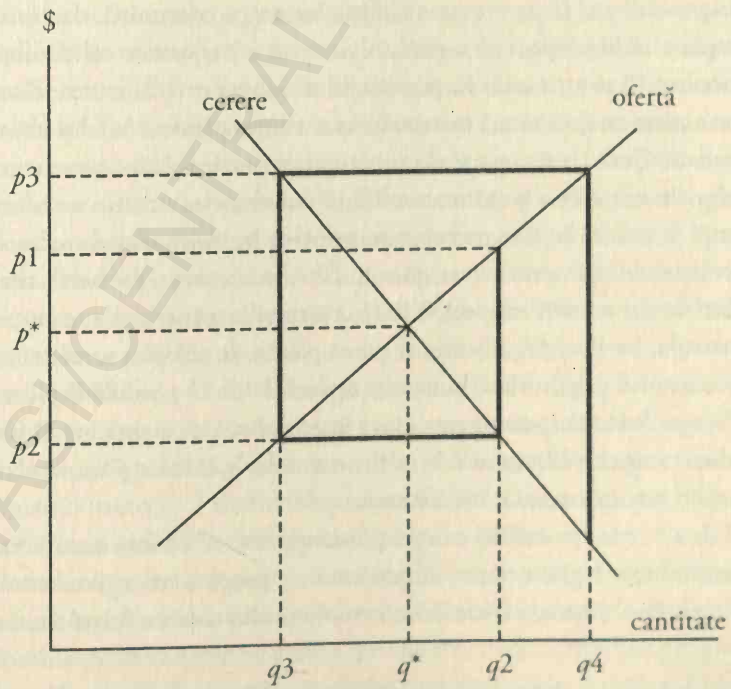

FIGURA 18.3

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

acel an. Prețul de clearing al pieței va fi $p3$, ceea ce îi determină pe fermieri să producă $q4$ în anul 4 și așa mai departe. În acest caz, prețurile și volumele formează o spirală spre exterior sau o formă de „pânză de păianjen", indicată de liniile îngroșate din diagramă. Surprizele plăcute alternează cu cele neplăcute, dar rezultatul așteptat nu se întâmplă niciodată. Dacă pantele relative ale cererii și ofertei se modifică, rezultatul ar putea fi o spirală spre interior, care converge spre prețul de echilibru p^* și volumul de echilibru q^*.

Există ceva irațional în legătură cu comportamentul fermierilor. Fiecare dintre ei crede că este liber să varieze producția pentru a-și maximiza profiturile, presupunând tacit în același timp că ceilalți fac în mod mecanic exact ce au făcut anul trecut. Cu toate că este probabil irațional, comportamentul este în mod evident inteligibil. Un filosof francez, Maurice Merleau-Ponty, a spus că tendința noastră spontană este de a-i vedea pe ceilalți drept „frați mai mici"[1]. Nu le atribuim cu ușurință celorlalți aceeași capacitate de reflecție și deliberare pe care introspecția ne spune că o avem noi înșine, la fel cum nu le atribuim nici aceleași zbateri interioare, îndoieli și angoase (vezi capitolul 23). Ideea de a-i vedea pe ceilalți calculați și tenace, așa cum ne vedem pe noi înșine, nu pare a ne fi foarte la îndemână.

Trei exemple preluate din comportamentul la vot pot ilustra această idee. Să presupunem că sunt un membru al aripii stângi a Partidului Socialist din țara mea. Aș prefera desigur să îi văd la putere pe socialiști, iar nu pe comuniști, dar, întrucât sondajele de opinie indică faptul că socialiștii vor avea o majoritate solidă după alegeri, votez cu comuniștii pentru a-mi forța partidul să adopte o viziune mai stângistă. Însă în niciun moment nu stau să mă întreb dacă nu cumva gândesc la fel și alți votanți cu viziune similară. Dacă o mare parte dintre ei au o viziune asemănătoare, comuniștii ar putea câștiga. Intenția de a produce rezultatul favorizat (o victorie socialistă dublată de o prezență puternică în favoarea comuniștilor) ar putea genera al treilea rezultat în ordinea preferințelor (o victorie comunistă). Într-un scenariu, probabil, ceva mai des întâlnit, dacă mulți votanți stau acasă în ziua alegerilor pentru că sunt încrezători în faptul că partidul lor va câștiga, acesta ar putea pierde. În cele din urmă, să ne aducem aminte de exemplul privitor la Chirac din capitolul 12. O posibilă explicație pentru solicitarea alegerilor anticipate poate consta în eșecul său de a anticipa că votanții vor trage concluzii asupra credințelor sale ca urmare a deciziei luate și nu vor crede pur și simplu în mod mecanic ceea ce indicau sondajele.

Eșecul de a-i vedea pe ceilalți ca agenți intenționați care caută maximizarea se observă în cazul în care legiuitorii sau administratorii propun strategii care sunt subminate când agenții se ajustează la ele. Conform dreptului roman, furtul unui singur cal

---

[1] Expresia acestuia este, de fapt, *younger brother* („frate mai mic"). [În textul lui Elster, apare *sibling*, termen care se referă atât la frați, cât și la surori. (N. trad.)]

sau bou transformă un om într-un hoț de vite, pe când nu ar fi o infracțiune dacă ar fura mai puțin de patru porci sau zece oi. Un comentator al legii a scris că, „într-un asemenea scenariu, ar trebui să ne așteptăm ca furturile de trei porci și opt oi să devină anormal de obișnuite". Pentru a promova siguranța locurilor de muncă, multe state au adoptat o legislație conform căreia este ilegal să disponibilizezi muncitori care au fost angajați, să zicem, de mai mult de doi ani sau mai mult. Angajatorii răspund în mod rațional prin *outsourcing* sau propunându-le angajaților contracte temporare, reducând astfel securitatea locului de muncă. Orașele pot construi autostrăzi pentru a reduce blocajele în trafic doar pentru a descoperi ulterior că, din moment ce mai mulți oameni merg acum cu mașina personală la locul de muncă, șoselele sunt la fel de blocate ca înainte, iar poluarea este mai mare. Un guvern poate încerca să limiteze imigrația persoanelor căsătorite cu cineva care se află deja cu forme legale în țară, efectul fiind acela că oamenii ajung să se căsătorească tocmai în acest scop. Scutirea studenților de recrutare creează un stimulent pentru a face o facultate. Chiar stabilirea explicită a unor sisteme de stimulente poate avea recul. Astfel, când sistemul public de educație din Chicago a creat, în 1996, un sistem prin care școlile neperformante erau pedepsite în diverse moduri, cei care au plănuit sistemul nu s-au gândit vreo clipă că, pe lângă crearea unui stimulent pentru profesori să muncească mai cu sârg, sistemul îi încuraja, de asemenea, să umfle notele, lucru care s-a și întâmplat.

Sindromul fratelui mai mic poate avea consecințe sociale importante, după cum se poate vedea în următoarele două exemple. Tocqueville observă că, în deceniile care au precedat Revoluția Franceză, clasele superioare au denunțat public viciile regimului și impactul său devastator asupra oamenilor, ca și cum cei din urmă ar fi fost surzi și nu ar fi auzit nimic: „Această chestiune îmi aduce aminte de sentimentul doamnei Duchâtelet care, conform secretarului lui Voltaire, nu se sfia să se dezbrace în fața servitorilor de sex masculin, nefiind convinsă că valeții erau și ei bărbați." În memoriile sale, secretarul nota chiar că „marile doamne își considerau lacheii drept simple automate". Afișarea simultană a disprețului față de clasele inferioare și denunțarea sărăciei din interior a pregătit spiritele pentru Revoluție. Mai recent, argumentul din spatele „curbei Phillips", conform căruia guvernul poate alege, dacă așa dorește, să scadă șomajul cu costul unei inflații mai mari, presupune că actorii sociali nu sunt conștienți de această politică. Când guvernele au încercat să atingă acest scop, însă, comportamentul strategic al sindicatelor raționale și al altor actori le-a subminat eforturile și a produs „stagflație" – inflație mare *și* șomaj ridicat.

Spre deosebire de consecințele neintenționate produse de externalități, cele generate de mecanismul fratelui mai mic pot lua sfârșit când agenții le înțeleg. Nu există strategii dominante în cazurile pe care le-am descris, ci doar strategii optime pornind de la ipoteza (de obicei, implicită) că ceilalți sunt mai puțin raționali. Odată ce toți agenții se percep drept raționali, comportamentul lor poate converge într-un rezultat

pe deplin predictibil. Toți crescătorii de porcine se așteaptă ca prețul de echilibru să predomine. Acționând conform acestor așteptări, ei vor produce volumul de echilibru. Credința lor împărtășită va aduce rezultatul pe care credința însăși îl cere. Această idee constituie subiectul capitolului următor.

◈◈◈

Notă bibliografică

Impactul dependenței asupra deprecierii timpului este documentat de L. Gordano *et al.*, „Mild opioid deprivation increases the degree that opioid-dependent outpatients discount delayed heroin and money", *Psychopharmacology*, 63 (2002), pp. 174-182. Observația care oferă un model al poveștii lui Andersen este inspirată de C.C von Weiszäcker, „Notes on endogenous change of tastes", *Journal of Economic Theory*, 3 (1971), pp. 345-372. Pentru o discuție despre poziția lui Marx cu privire la consecințele neintenționate, vezi cartea mea *Making Sense of Marx* (Cambridge University Press, 1985), capitolul 1.3.2 și *passim*. Modelul dependenței își are originea în G. Becker și K. Murphy, „A theory of rational addiction", *Journal of Political Economy*, 96 (1988), pp. 675-700. Un studiu conceptual superb al consecințelor neintenționate îl reprezintă cartea lui T. Schelling, *Micromotives and Macrobehavior* (New York, Norton, 1978). Pentru ideea de internalități, vezi R. Hernstein *et al.*, „Utility maximization and melioration: Internalities in individual choice", *Journal of Behavioral Decision Making*, 6 (1993), pp. 149-185. Exemplul cu privire la sistemul public de educație din Chicago e luat din B. Jacob și S. Leavitt, „Rotten apples: An investigation of the prevalence and predictors of teacher cheating", *Quarterly Journal of Economics*, 118 (2003), pp. 843-877.

BCU IASI CENTRAL UNIVERSITY LIBRARY

Capitolul 19

INTERACȚIUNEA STRATEGICĂ

Acțiunea strategică și alegerile simultane

Inventarea teoriei jocurilor poate ajunge să fie considerată drept cea mai importantă descoperire a științelor sociale din secolul XX. Teoria este valoroasă în parte pentru caracterul său explicativ, dar mai cu seamă din punct de vedere conceptual. În anumite cazuri, ne permite să explicăm comportamente care, înainte, păreau paradoxale. Mai important, face lumină asupra structurii interacțiunii sociale. Odată ce vezi lumea prin lentilele teoriei jocurilor – sau „teoria deciziilor interdependente", cum ar mai putea fi numită –, nimic nu mai pare la fel.

Voi lua mai întâi în considerare jocuri în care agenții iau decizii simultan. Scopul este acela de a înțelege dacă și cum n agenți sau *jucători* pot realiza o coordonare neimpusă a *strategiilor* lor. De multe ori, vom aborda cazul special în care $n = 2$. Jucătorii pot fi în măsură să comunice unii cu alții, dar să nu intre în înțelegeri care presupun un contract ferm. Fiecărui n-tuplu de strategii, câte una aleasă de fiecare agent, îi corespunde un *rezultat*. Fiecare agent clasifică rezultatele posibile în funcție de *ordinea* personală a *preferințelor*. Atunci când este nevoie, vom presupune că sunt satisfăcute condițiile pentru reprezentarea preferințelor sub forma utilităților cardinale (capitolul 11). *Structura recompensei* este funcția pe care fiecare n-tuplu de strategii o atribuie unui n-tuplu de utilități. Cu toate că termenul „recompensă" poate sugera un rezultat în termeni monetari, cuvântul va fi folosit pentru a ne referi la rezultate de natură psihologică (utilitățile sunt, în cele din urmă, preferințe). Când, după cum este adesea cazul, structura recompensei materiale sau monetare și structura recompensei psihologice sunt divergente, doar cea din urmă este relevantă.

După cum am menționat în trecere în ultimul capitol, un agent poate avea o strategie *dominantă* în sensul că, indiferent ce fac ceilalți, aceasta produce un rezultat mai bun pentru agent decât orice alt rezultat al alegerii unei strategii alternative. *Rezultatul* agentului poate depinde de ceea ce fac ceilalți, dar *alegerea* nu depinde. În alte cazuri, există o interdependență autentică a alegerilor. În cazul în care ceilalți conduc

pe partea stângă a drumului, cel mai bun răspuns pe care eu îl pot da este să mă conformez; în cazul în care ceilalți merg pe dreapta, cel mai bun răspuns pe care îl pot da este să merg pe dreapta.

Un *echilibru* este un *n*-tuplu de strategii cu proprietatea că niciun jucător nu poate, prin deviere de la strategia sa de echilibru, să determine unilateral producerea unui rezultat pe care îl preferă strict față de rezultatul de echilibru. Echivalent, în echilibru, strategia aleasă de fiecare jucător este cel mai bun răspuns la strategiile alese de ceilalți, în sensul slab că acesta nu poate face *ceva mai bun* decât să aleagă strategia sa de echilibru dacă ceilalți o aleg pe a lor. Însă strategia nu trebuie să fie optimă în sensul tare, în care agentului i-ar fi *mai rău* dacă ar devia unilateral. În general, un joc poate avea mai multe stări de echilibru. Vom vedea câteva exemple în curând. Să presupunem, însă, că există un singur echilibru. Să presupunem, de asemenea, că structura recompensei și raționalitatea tuturor jucătorilor sunt binecunoscute.[1] Presupunând acest lucru, putem prezice că toți agenții vor alege strategia lor de echilibru din moment ce este singura bazată pe credințe raționale despre ce vor face ceilalți.

Unele jocuri cu echilibru unic se bazează pe existența unor strategii dominante. Expresia „se bazează pe existența unor strategii dominante" poate însemna două lucruri, după cum se poate vedea în casetele A și B din figura 19.1.[2] Într-un accident în care au fost implicate două mașini, ambele suferă daune. Într-un accident în care au fost implicați un pieton și o mașină, doar pietonul este rănit. Accidentele mașină-mașină se produc dacă cel puțin unul dintre șoferi este neatent. Dacă ambii sunt neatenți, rezultatul va fi și mai rău. Accidentele mașină-pieton se produc doar dacă atât șoferul, cât și pietonul sunt neatenți. A fi atent este un lucru costisitor. Din aceste premise decurge că, în situația mașină-mașină, faptul de a fi atent este strategia dominantă pentru fiecare șofer. În situația mașină-pieton, neatenția este strategia dominantă pentru șofer. Pietonul nu are vreo strategie dominantă întrucât atenția

---

[1] Un fapt este binecunoscut dacă toată lumea îl știe, toată lumea știe că ceilalți îl știu, toată lumea știe că ceilalți știu că toată lumea știe și așa mai departe. Pentru a evita folosirea expresiei „și așa mai departe", care sugerează un șir infinit de credințe, ideea poate fi formulată și astfel: nu există *n*, astfel încât un fapt este binecunoscut până la nivelul *n* al șirului, dar nu la nivelul *n* + 1. Pentru o ilustrare simplă, acest tip de cunoaștere se poate realiza într-o sală de curs. Când profesorul le spune studenților un fapt, toți îl știu, toți știu că ceilalți îl știu și așa mai departe.

[2] Prin convenție, primul număr din fiecare celulă reprezintă recompensa pentru „jucătorul marcat pe rânduri", care alege între strategiile de deasupra și de dedesubt, iar a doua recompensă este pentru „jucătorul marcat pe coloană", care alege între strategiile din stânga și din dreapta. În funcție de context, recompensele pot fi utilități cardinale, utilități ordinale, bani sau orice altceva despre care se presupune că jucătorii vor să maximizeze. În Figura 19.1, putem considera că recompensele reprezintă utilități ordinale care reflectă preferințele cu privire la rezultate. Am adoptat convenția de a încercui echilibrele.

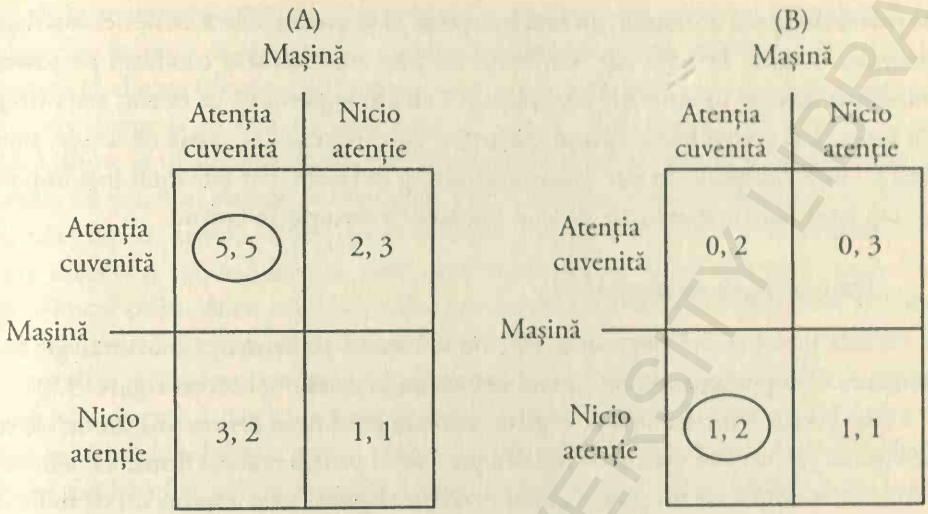

FIGURA 19.1

care se cuvine este cel mai bun răspuns pentru neatenție, iar neatenția este cel mai bun răspuns pentru atenția cuvenită. Întrucât știe că strategia dominantă a șoferului este neatenția și, pentru că este rațional, o va alege pe aceasta, pietonul va alege oricum strategia atenției cuvenite.[1]

Jocurile în care toți jucătorii au o strategie dominantă sunt destul de des întâlnite și importante din punct de vedere empiric, după cum vom vedea. Din punct de vedere teoretic, acestea sunt întrucâtva banale, cu excepția cazurilor în care sunt repetate în timp. Jocurile în care unii jucători au strategii dominante și pot determina alegeri clare pentru ceilalți sunt mai rare, dar sunt importante, de asemenea. Acestea, însă, impun condiții mai tari din punct de vedere informațional întrucât, în exemplul nostru, pietonul trebuie să cunoască rezultatele posibile atât pentru șofer, cât și pentru sine, pe când cei doi șoferi nu trebuie să cunoască decât rezultatul propriu. De multe ori, putem atribui strategii dominante celorlalți fără prea multe probleme. De exemplu, în mod obișnuit nu ne uităm în ambele direcții înainte să traversăm un o stradă cu sens unic deoarece presupunem că, din frica de a nu face vreun accident, șoferii vor respecta regula sensului unic.

O clasă specială de jocuri are *echilibre coordonate*, de multe ori numite „convenții", în care fiecare jucător nu numai că nu are niciun motiv să devieze unilateral, dar ar prefera ca nimeni altcineva să nu o facă. Într-un echilibru în care fiecare conduce

[1] Conform unor analize de drept, o funcție importantă a dreptului civil este de a folosi un sistem de amenzi și daune pentru a modifica matricea recompenselor, astfel încât echilibrele care rezultă au o anumită proprietate dezirabilă (eficiență sau echitate).

B.C.U. M. EMINESCU IAȘI

pe partea dreaptă a drumului, un accident poate să se producă dacă eu deviez sau dacă altcineva deviază. În acest caz, echilibrul nu este unic întrucât condusul pe partea stângă are aceleași proprietăți.[1] De multe ori nu are importanță ce facem, atâta timp cât facem toți același lucru. Sensul cuvintelor este arbitrar, dar, odată ce acestea sunt fixate, devin convenții. În alte cazuri, contează și ce facem, dar este mult mai important să facem toți același lucru. Mă voi întoarce cu exemple în curând.

Două exemple de duopol

Unele jocuri au echilibre unice care nu se bazează pe existența unor strategii dominante. Comportamentul de duopol este un astfel de exemplu (vezi Figura19.2).

Când două firme domină pe o piață, scăderea producției pentru una dintre ele va determina prețuri mai mari și o creșterea producției pentru cealaltă firmă. Cu alte cuvinte, fiecare firmă are un plan al „celui mai bun răspuns" care îi spune cât de mult să crească producția în funcție de producția celeilalte firme. La echilibru, producția fiecărei firme este cel mai bun răspuns la producția celeilalte. Acest enunț nu înseamnă că firmele nu ar putea să fie într-o situație mai bună. Dacă formează un cartel și își restrâng producția sub nivelurile de echilibru, ambele pot obține profituri mai mari. Însă aceste niveluri optime de producție nu sunt cele mai bune răspunsuri față de celălalt. Firmele se confruntă, de fapt, cu o dilemă a prizonierului (definită în Figura 19.3).

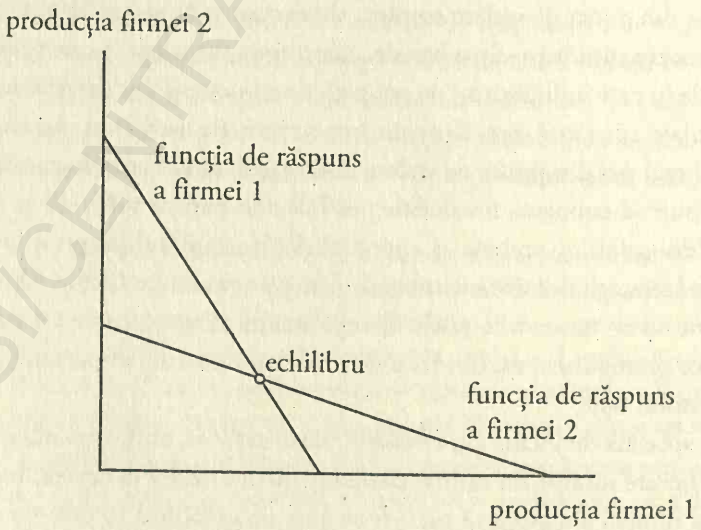

FIGURA 19.2

[1] Deși faptul că nu este unic nu decurge din definiția formală, aceasta pare a fi trăsătura generală a jocurilor de coordonare din viața reală.

Pentru alt caz de duopol, putem lua în considerare doi vânzători de înghețată pe plajă care încearcă să găsească cea mai bună amplasare pentru standuri pornind de la ipoteza că clienții (despre care presupunem că sunt egal distribuiți de-a lungul țărmului) vor merge la cel mai apropiat stand. Nu există nicio strategie dominantă. Dacă unul dintre ei își amplasează standul la o anumită distanță în stânga față de mijlocul plajei, cel mai bun răspuns al celuilalt ar fi să își amplaseze standul simetric, în partea dreaptă față de centru, la care cel mai bun răspuns al primului va fi să se mute ușor spre dreapta și așa mai departe, până când standurile lor se vor afla unul lângă altul la mijlocul plajei. Acest echilibru unic nu este, evident, cel mai bun pentru totalitatea clienților. Pentru ei, cel mai bun rezultat ar fi acela în care un stand este situat în punctul de mijloc între centrul plajei și capătul ei, iar celălalt stand simetric, de partea cealaltă a centrului plajei. Cu toate că acest rezultat este la fel de bun pentru vânzători ca rezultatul de echilibru, aceste poziții nu sunt cel mai bun răspuns față de celălalt. Acest model a fost aplicat și pentru a explica tendința partidelor politice (într-un sistem bipartit) de a se deplasa spre centrul spectrului politic.

Să presupunem, însă, că atunci când ambele standuri se află la mijloc, clienții cei mai apropiați de capătul plajei se vor abține să mai cumpere înghețată pentru că ar însemna să se topească până se vor întoarce la locul lor pe plajă. Dacă niciun client nu e dispus să meargă mai mult de jumătate din lungimea plajei, un sfert pentru a ajunge la stand și alt sfert pentru a ajunge înapoi, rezultatul optim al consumatorilor este, de asemenea, echilibrul unic, din moment ce niciunul nu este stimulat în vreun fel să se mute. Să presupunem că plaja are 1 000 de metri. Dacă vânzătorul de la 750 de metri se mută cu standul la 700, acesta va pierde 50 de clienți situați între 950 și 1 000 de metri, care nu sunt dispuși să meargă mai mult de 450 de metri și va câștiga 25 de clienți între 475 și 500 de metri pentru care standul este mai aproape decât celălalt – o pierdere netă. Un argument asemănător ar putea fi folosit pentru a explica de ce partidele politice nu converg pe deplin spre centru, întrucât extremiștii de la fiecare capăt al spectrului ar prefera să se abțină decât să voteze pentru un partid centrist. Pe lângă aceste aspecte, după cum am observat la sfârșitul capitolului 17, pur și simplu nu este plauzibil să vezi maximizarea votului drept singurul scop al partidelor politice.

Câteva jocuri frecvente

Câteva structuri simple de interacțiune, cu recompense ca în Figura 19.3 se produc destul de des într-un număr mare de contexte.[1] C și D înseamnă „cooperare" și, respectiv, „trădare". În Jocul Telefonului, jucătorul de pe coloană este primul sunat. În Jocul Punctului Focal, A și B pot fi orice pereche de acțiuni, astfel încât ambii jucători

[1] Cu toate că sunt enunțate aici ca jocuri de două persoane, pot fi generalizate ușor pentru a include mai mulți agenți. O versiune de $n+1$ persoane a Dilemei Prizonierului este ilustrată, de exemplu, în Figura 24.2.

ar prefera să se coordoneze de fiecare decât să nu se coordoneze, dar sunt indiferenți între cele două moduri în care se pot coordona.

Jocurile pun în evidență structura a două chestiuni esențiale în interacțiunea socială – *cooperarea* și *coordonarea*. Într-o societate în care nu există cooperare în vederea beneficiului reciproc, viața ar fi „singuratică, săracă, periculoasă, sălbatică și scurtă" (Hobbes). Că ar fi *predictibil* de rea este o palidă consolare. Într-o societate în care oamenii nu sunt capabili să își coordoneze comportamentul, consecințele neintenționate ar fi atât de abundente, încât viața ar fi ca „un basm de furii și de nerozie, băznit de-un prost și făr' de nicio noimă" (*Macbeth*). Atât cooperarea, cât și coordonarea au uneori succes, însă de multe ori se întâmplă să eșueze dramatic. Teoria jocurilor poate face lumină atât asupra succeselor, cât și asupra eșecurilor.

Dilema Prizonierului (DP), Vânătoarea de Cerbi și Lașul implică în mai multe moduri alegerea între cooperare și trădare (necooperare). Dilema Prizonierului se numește astfel datorită următoarei istorisiri folosite într-o discuție anterioară pentru a o pune în evidență. Fiecăruia dintre doi prizonieri care au fost complici la aceeași infracțiune, dar care sunt ținuți în celule diferite, i se spune că, dacă îl toarnă pe celălalt, dar celălalt nu îl toarnă, va fi eliberat, iar celălalt va merge la pușcărie zece ani; dacă niciunul nu îl toarnă pe celălalt, ambii vor executa o pedeapsă de un an, iar dacă ambii se toarnă unul pe celălalt, ambii vor sta cinci ani la închisoare.[1] În aceste circumstanțe, strategia dominantă este să îl torni pe celălalt, cu toate că amândurora le-ar fi mai bine dacă nu ar face-o. Rezultatul este generat de o combinație între „tentația actelor parazitare" (libertatea) și „frica de a nu fi fraierit" (executarea a zece ani).

Externalitățile negative discutate în ultimul capitol pot fi privite din perspectiva unor DP implicând mai multe persoane. Urmează unele exemple. Pentru fiecare muncitor (presupunând motivații egoiste) este mai bine să nu facă parte dintr-un sindicat decât să facă parte, chiar dacă este mai bine pentru toți dacă toți fac parte dintr-un sindicat și câștigă mai mult. Pentru fiecare firmă dintr-un cartel este mai bine să pătrundă pe piață și să producă la un volum mai mare pentru a specula prețurile mari determinate de restricțiile impuse asupra productivității celorlalte firme, dar, dacă toate vor face la fel, prețurile vor scădea la nivelul lor competitiv; maximizarea profitului de către fiecare firmă subminează maximizarea profiturilor considerate împreună. Organizația Țărilor Exportatoare de Petrol (OPEC), un cartel, este vulnerabilă în același mod. Alte exemple sunt situațiile în care toată lumea trebuie să fugă cât poate de repede pentru a rămâne în același loc, cum ar fi cursa înarmării dintre Statele

[1] Recompensele pentru Dilema Prizonierului din Figura 19.3 ar putea părea artificiale. Pentru scopurile de față, tot ce contează este clasificarea (ordinală) a rezultatelor. Mai târziu, recompensele vor fi reinterpretate ca recompense bănești.

	C	D
C	3, 3	0, 4
D	4, 0	1,5, 1,5

Dilema Prizonierului

	C	D
C	4, 4	1, 3
D	3, 1	2, 2

Vânătoarea de Cerbi/Jocul Asigurării

	C	D
C	2, 2	1, 3
D	3, 1	0, 0

Laşul

	Balet	Box
Balet	1, 2	0, 0
Box	0, 0	2, 1

Bătălia Sexelor

	Formează din nou	Nu forma
Nu forma	2, 2	0, 0
Formează din nou	0, 0	1, 1

Jocul Telefonului

	A	B
A	1, 1	0, 0
B	0, 0	1, 1

Jocul Punctului Focal

FIGURA 19.3

BCU INSTITUTUL CENTRAL UNIVERSITY LIBRARY

Unite și Uniunea Sovietică, publicitatea politică sau studenții care scriu lucrări pentru un profesor care „dă note în funcție de curba lui Gauss".

Ideea din spatele Vânătorii de Cerbi este atribuită lui Jean-Jacques Rousseau, cu toate că limbajul acestuia este uneori opac.[1] Într-o formă mai stilizată, acesta presupune doi vânători care pot alege între a vâna cerbi (C) și a vâna iepuri (D). Fiecare poate prinde un iepure pe cont propriu, dar, pentru a vâna un cerb, este necesar (și suficient) ca eforturile lor să fie conjugate. O jumătate de cerb este mult mai valoroasă decât un iepure. E nevoie de mult mai mult timp și efort pentru a vâna iepuri dacă ambii fac acest lucru pentru că zgomotul făcut de celălalt vânător sperie prada. La fel ca în Dilema Prizonierului, există riscul de a fi fraierit, dacă unul vânează cerbul, pe când celălalt urmărește iepurele. Însă nu există și tentația actelor parazitare. Jocul are două stări de echilibru, în celula din stânga sus și în celula din dreapta jos.

Cu toate că primul echilibru este în mod evident mai bun, el nu poate fi realizat. Pentru a vedea de ce, putem renunța la presupoziția că structura recompensei este binecunoscută și putem îngădui agenților să aibă credințe eronate cu privire la structura de recompensă a celorlalți agenți. Acțiunile întreprinse pe baza acestor credințe vor forma un *echilibru în sens slab* dacă, pentru fiecare agent, acțiunile întreprinse de ceilalți le confirmă credințele despre ei. Să presupunem, că, într-o Vânătoare de Cerbi, fiecare agent crede în mod eronat că ceilalți au preferințe de tip DP. Dată fiind această credință, acțiunea rațională este aceea de a trăda, confirmând astfel credința că ceilalți au preferințe de tip DP. O astfel de societate poate sfârși cu niveluri mari de evaziune fiscală și corupție. Mă voi întoarce la astfel de cazuri de „ignoranță pluralistă" în capitolul 23. Într-o altă societate, unde oamenii cred în mod corect că ceilalți au preferințe de tip Vânătoare de Cerbi, va apărea un echilibru bun în care oamenii își plătesc taxele și nu oferă sau iau mită. „Cultura corupției" s-ar putea să fie un fenomen dependent de credințe, și nu de motivație.

Controlul la nivel internațional al bolilor infecțioase are o structură de tip Vânătoare de Cerbi. Dacă o singură țară va eșua în implementarea măsurilor necesare, celelalte nu vor fi în măsură să se protejeze.[2] Putem lua în considerare un alt exemplu: măsurile antiteroriste. Dacă doar una dintre cele două națiuni va investi în astfel de măsuri, beneficiile se vor răsfrânge și asupra celeilalte. În cazul în care costurile vor depăși beneficiile pentru sine, țara respectivă nu va investi unilateral. Însă, dacă

---

[1] „Pentru a prinde un cerb, fiecare observă că, pentru a avea succes, trebuie să își facă datoria: dar dacă un iepure se întâmplă să fie mai ușor de prins pentru vreunul dintre ei, nu încape îndoială că acesta trebuie urmărit fără scrupule și, odată prada asigurată, nu îi mai pasă că, prin aceasta, a provocat ratarea celuilalt." Am putea înțelege de aici că urmărirea iepurilor este strategia dominantă.

[2] Aceasta este o simplificare enormă, făcută doar de dragul exemplificării.

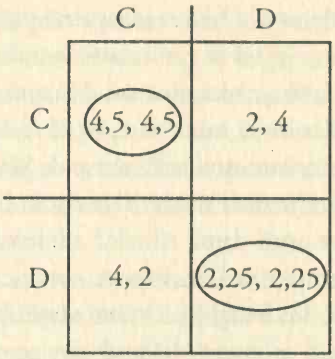

FIGURA 19.4

ambele țări investesc, abilitatea de a obține informații ar putea duce la un nivel mai mare de securitate pentru fiecare decât acela obținut din exploatarea investiției făcută de una singură

În aceste exemple, structura recompensei reiese din natura cauzală a situației. În Vânătoarea de Cerbi și în cazul controlului bolilor, „tehnologia de prag" implică faptul că eforturile individuale nu au niciun rost. În cazul antiterorismului, cauza subiacentă este ceva de felul economiilor de scală: zece unități de efort au un efect mai mult decât dublu comparativ cu cinci unități. În alte cazuri, structura recompensei se datorează faptului că agenților le pasă de alte lucruri decât de recompensele lor materiale. În astfel de cazuri, ne putem referi la joc ca un Joc al Asigurării (JA). Chiar dacă structura recompenselor materiale este cea a unei DP, fiecare individ poate fi dispus să coopereze dacă este asigurat de faptul că și ceilalți o vor face. Dorința de a fi corect sau reticența de a fi parazitar pot fi mai mari decât tentația de a exploata cooperarea cu ceilalți. În schimb, preferințele altruiste pot transforma o DP într-un JA.

Să interpretăm recompensele din DP prezentate în Figura 19.3 ca recompense bănești și să presupunem că utilitatea fiecărei persoane este egală cu recompensa monetară plus jumătate din recompensa monetară a celuilalt. În acest caz, recompensa de utilitate va fi ca în Figura 19.4, adică un JA. DP poate fi transformată într-un JA printr-un al treilea mecanism, dacă cineva din exterior impune o penalizare pentru alegerea strategiei necooperante D. Dacă interpretăm din nou recompensele DP din Figura 19.3 ca recompense monetare *și* presupunem că agenților le pasă doar de acest lucru, faptul de a deduce 1,25 din recompensa pentru trădare o va transforma într-un JA. De exemplu, un sindicat ar putea să impună sancțiuni formale sau informale asupra muncitorilor care nu sunt reprezentați de un sindicat. În cele din urmă, am putea transforma DP într-un AG recompensând cooperarea și oferind, de exemplu, un bonus sau o mită de 1,25 pentru aceia care cooperează. Promisiunile unor recompense trebuie însă respectate, în timp ce o amenințare nu trebuie să fie pusă în practică dacă

lucrurile merg așa cum se dorește. Dacă recompensa parazitului este foarte mare, beneficiile aduse de cooperare ar putea să nu fie suficient de mari pentru a finanța mita.[1] În anumite cazuri, însă, se folosesc recompense. Muncitorii care intră într-un sindicat pot beneficia nu doar de salarii mai mari, care, de obicei, cresc în același ritm cu salariile muncitorilor fără reprezentare sindicală, dar și de planuri de pensionare mai bune și de vacanțe mai ieftine, oferite doar membrilor de sindicat.

Jocul Lașului a fost denumit după ritualul adolescenților din *Rebel fără cauză*, un film din 1955. Adolescenții din Los Angeles conduc mașini furate pe vârful unei stânci și joacă un joc în care doi băieți își conduc mașinile foarte aproape de buza prăpastiei oprindu-se în ultimul moment. Băiatul care se oprește primul este „lașul" și pierde jocul. Într-o altă versiune, doi șoferi îndreaptă mașinile una spre cealaltă, iar șoferul care virează primul este „lașul". În fiecare dintre cele două echilibre, fiecare agent procedează invers decât celălalt. Chiar dacă structura recompensei și raționalitatea agenților sunt binecunoscute, nu putem prezice care dintre echilibre va fi ales (dacă este ales vreunul). Din punctul de vedere al alegerii raționale, situația este *nedeterminată*. În a doua versiune („virajul") a jocului, un jucător poate încerca să evite nedeterminarea legându-se la ochi (celălalt șofer vede acest lucru), determinându-l pe celălalt să vireze. Însă și aici avem aceeași situație, în care opțiunile „legat la ochi" și „nelegat la ochi" înlocuiesc „virează" și „nu virează".[2] Este o situație foarte frustrantă.

O interpretare a cursei înarmării este aceea că are structura Lașului. Criza rachetelor cubaneze este citată adesea drept caz în care două superputeri au ajuns să fie blocate într-o confruntare de tip Lașul, iar URSS a fost prima care să se „lege la ochi". Un alt exemplu presupune doi fermieri care folosesc același sistem de irigație pentru a-și uda terenurile. Sistemul poate fi întreținut de o singură persoană în mod adecvat, dar ambii fermieri au de câștigat în mod egal de pe urma lui. Dacă un fermier nu se achită de datoria sa de întreținere, poate fi încă în interesul celuilalt fermier să o facă. Cazul lui Kitty Genovese poate fi privit, de asemenea, dintr-o astfel de perspectivă dacă presupunem că fiecare vecin ar prefera să intervină dacă și numai dacă nimeni altcineva nu o face.

Să ne întoarcem acum la chestiunea *coordonării*, luând în considerare mai întâi Bătălia Sexelor. Stereotipul din spatele poveștii este următorul. Un bărbat și soția sa vor să iasă în oraș într-o seară. Aceștia sunt decișI să meargă după serviciu fie la un spectacol de balet, fie la un meci de box și discută decizia finală prin telefon. Însă

---

[1] Indiferent dacă sunt folosite pedepsele sau recompensele, costurile implementării unui sistem și ale monitorizării agenților trebuie finanțate din câștigurile produse de cooperare. Mai ales din acest motiv, astfel de aranjamente pot fi imposibile sau pot presupune risipă.

[2] Asemănător, „soluția" la Dilema Prizonierului, care constă în promisiunea fiecăruia de a coopera, nu face decât să recreeze DP în care alegerile noi sunt „mă țin de promisiune" și „nu mă țin de promisiune".

telefonul bărbatului se strică, prin urmare trebuie să decidă printr-o coordonare tacită. Aceștia au un interes comun: acela de a fi împreună, dar au interese divergente despre locul în care să meargă. La fel ca jocul Lașului, și acesta are două echilibre, coordonarea în privința baletului sau a meciului de box. Și, la fel ca în jocul Lașului, nici caracterul binecunoscut al structurii recompensei și nici raționalitatea nu le vor fi de folos celor doi pentru a le spune unde să se întâlnească. Din nou, situația este indeterminată.

Jocurile de acest tip pot apărea când coordonarea ia mai multe forme, care sunt toate mai bune pentru agenți decât lipsa completă a coordonării, dar fiecare este preferat de anumiți agenți mai mult decât de ceilalți.[1] În viața socială și politică, aceasta este mai degrabă regula decât excepția. Toți cetățenii pot prefera orice formă de constituție politică (ale cărei regimuri politice se mențin între anumite limite) decât lipsa completă a unei constituții, având în vedere că stabilitatea pe termen lung este importantă pentru capacitatea fiecăruia de a-și face planuri. Când legea este stabilă și greu de schimbat, comportamentul poate fi reglementat în funcție de aceasta. Însă fiecare grup de interese poate prefera o constituție specifică în limitele date: creditorii fac lobby pentru interzicerea banilor de hârtie prin constituție, fiecare partid politic militează pentru un sistem electoral care să îi fie favorabil, cei care au un candidat puternic la președinție vor sporirea prerogativelor acelei instituții și așa mai departe.

Echilibrele coordonării multiple apar când societăți diferite dezvoltă inițial standarde diferite de greutate, lungime și volum pentru a descoperi mai târziu beneficiile unei soluții comune. Europa continentală și lumea anglo-saxonă rețin standarde diferite în aceste zone. Spre deosebire de soluțiile constituționale multiple, obstacolul în calea unei înțelegeri nu este o divergență permanentă de interese, ci doar costuri de tranziție de scurtă durată. Alegerea unui standard poate fi, însă, un joc al Lașului. Să presupunem, în mod implauzibil, că standardul este scris în constituție ca o clauză protejată (imună la oricare amendament). Fiecare țară va fi stimulată să adere la acest standard înaintea celorlalte.

Jocul Telefonului este definit de nevoia unei reguli care să le spună părților ce să facă dacă o conversație telefonică se întrerupe din greșeală. Există două echilibre de coordonare: numărul trebuie format din nou de către persoana care a dat telefonul sau de persoana care l-a primit. Fiecare regulă este mai bună decât să nu formeze nimeni sau să formeze ambii. Spre deosebire de Bătălia Sexelor, în acest caz există un echilibru mai bun pentru ambii decât celălalt. Este mult mai eficient pentru acela care a dat telefon la început să formeze din nou întrucât acesta are cele mai mari șanse de a ști la ce număr să sune. Agenții raționali pe deplin informați vor converge spre un echilibru de

[1] După cum vom vedea mai târziu (capitolul 25), această problemă a împărțirii beneficiilor obținute din cooperare poate fi studiată și în teoria negocierii, o ramură mai specializată a teoriei jocurilor.

coordonare superior. Acest enunț ignoră, însă, costurile presupuse de reformarea numărului. În cazul în care costurile sunt mari, jocul devine Bătălia Sexelor.

În cele din urmă, să luăm în considerare Jocul Punctului Focal, care poate fi ilustrat de o versiune a Bătăliei Sexelor. Soții s-au înțeles să se uite la un film care rulează și în cinematograful A, și în cinematograful B, dar au amânat alegerea exactă a locului în care vor merge. Presupunem că niciunul dintre cinematografe nu este mai apropiat sau mai convenabil decât celălalt. La fel ca în Bătălia Sexelor, informația, rațională tatea și lucrurile binecunoscute în sine nu le vor putea sugera unde să meargă. Însă ar putea exista un indiciu psihologic în această situație, care va funcționa ca „punct focal" al coordonării. Dacă prima lor întâlnire de cuplu a avut loc în cinematograful A, acest lucru îi poate face să conveargă spre acel loc. În acest caz, indiciul este un eveniment privat. În alte cazuri, indiciile pot fi împărtășite de o populație mai mare. Printre newyorkezi, de exemplu, tradiția spune că, dacă dacă vrei să te întâlnești cu cineva, ar trebui să o faci la amiază la ceasul cel mare din Gara Grand Central. Și, chiar dacă nu ar exista vreo uzanță, mulți oameni s-ar întâlni la gară întrucât, în multe orașe, gara este cea mai importantă clădire a orașului.[1] Unicitatea sa o transformă într-un punct focal. Miezul zilei are aceeași proprietate.[2]

Acest efect de focalizare este ușor de demonstrat prin experimente. Dacă membrii unui grup sunt rugați să scrie un număr întreg pozitiv pe o foaie de hârtie și li se spune că toți vor primi o recompensă dacă scriu același număr, aceștia vor converge invariabil spre 1. Există un unic număr care este cel mai mic număr întreg, dar nu există un unic număr care este cel mai mare număr întreg. În alte contexte, 0 ar putea să fie punctul focal. În dezbaterile din timpul Războiului Rece cu privire la utilizarea armelor nucleare tactice de către Statele Unite, fără escaladarea unui război nuclear propriu-zis, au fost sugerate mai multe idei pentru un „standard clar" care să permită utilizarea lor limitată. Într-un final, s-a decis că punctul focal este acela de a *nu folosi* vreo armă.

Pascal face o observație asemănătoare cu privire la importanța obiceiurilor: „De ce respectăm legile vechi și opiniile din vechime? Oare pentru că sunt mai bune? Nu, ci pentru că sunt unice și elimină sursele diversității." Altundeva scrie:

> Cele mai nerezonabile lucruri din lume devin cele mai rezonabile pentru că oamenii sunt atât de dezechilibrați. Ce ar putea fi mai puțin rezonabil decât a-l alege drept conducător al unui stat pe fiul cel mai în vârstă al reginei? Nu-l alegem căpitan al unei nave pe cel mai în vârstă dintre marinari. O astfel de lege ar fi

---

[1] În New York, cei care nu cunosc tradiția nu se pot duce în Gara Grand Central pentru că nu este singura gară (mai există și Penn Station). S-ar putea, însă, coordona în funcție de Empire State Building.

[2] Cu toate că și miezul nopții este un punct focal, este inferior miezului zilei, fiind mult mai puțin convenabil.

ridicolă și nedreaptă, dar, pentru că oamenii sunt oameni și vor fi întotdeauna la fel ca acum, acest lucru devine rezonabil și drept. Cine altcineva să fie ales? Omul cel mai virtuos și mai abil? Deja am fi la cuțite cu toți aceia care susțin că ei sunt cei mai virtuoși și mai abili. Să atașăm, deci, această caracterizare unui aspect necontroversat. El este cel mai în vârstă fiu al regelui: măcar atâta lucru este clar, aici nu mai încape loc de dispută. Rațiunea nu poate mai mult în acest caz, întrucât războiul civil este cel mai mare dintre rele.

Raționamentul acesta poate influența alegerea unui rege când există mai mulți pretendenți la tron. Când s-a ales regele Franței în timpul Restaurării, Talleyrand a argumentat cu succes că moștenitorul legitim al ultimului rege al Franței era unicul punct focal care ar fi putut să prevină conflictele. După cum scrie în memoriile sale, „un rege impus ar fi rezultatul forței sau intrigii; nici una, nici alta nu este suficientă. Pentru a întemeia un sistem durabil care să fie acceptat fără opoziție, trebuie acționat în baza unui principiu." Mai târziu, Marx a susținut că Republica din 1848 și-a datorat existența faptului că era cea de-a doua opțiune în ordinea preferințelor pentru fiecare dintre cele două ramuri ale familiei regale. Tocqueville face o observație asemănătoare pentru a explica stabilitatea domniei lui Napoleon al III-lea. Și democrația poate fi văzută ca o soluție de tip punct focal. Când există mai multe temeiuri calitative competitoare pornind de la care oamenii pot avea pretenții de superioritate – înțelepciune, bogăție, virtute, naștere –, soluția cantitativă a regulii majorității capătă o importanță unică. Fostele state coloniale, în care triburile vorbeau limbi diferite, pot alege limba colonizatorului pentru scopuri oficiale. Părțile aflate în litigiu converg ușor spre o alternativă care este a doua în ordinea preferințelor pentru toate părțile implicate.

În 1989, reînhumarea lui Imre Nagy a provocat un punct focal pentru ca 250 000 de oameni să mărșăluiască pe străzile Budapestei semnalându-și nemulțumirea față de regim. Ca și în exemplele precedente, punctul focal a permis cooperarea prin coordonare. În situații de conflict, punctele focale pot avea efecte destul de diferite. În Războiul din Crimeea, generalul francez Pélissier a luat decizia de a fixa al doilea atac asupra Sevastopolului pe 18 iunie 1855 pentru că voise să îi facă pe plac lui Napoleon al III-lea câștigând o victorie în ziua aniversării Bătăliei de la Waterloo. Pentru că importanța acestei date era binecunoscută pentru ruși, aceștia au fost în măsură să anticipeze și l-au înfrânt.

O lecție care poate fi desprinsă din această trecere în revistă este că o situație dată din lumea reală poate fi modelată prin apel la mai multe jocuri, totul depinzând de presupozițiile suplimentare. Cursa înarmării poate fi modelată ca o DP, ca Lașul sau ca JA. Afilierea la un sindicat poate fi văzută ca o DP sau ca un JA. Formarea unui număr din nou poate fi văzută ca o Bătălie a Sexelor sau ca un Joc al Telefonului. Coordonarea greutăților și măsurilor poate fi văzută ca un joc al Lașului sau ca o

Bătălie a Sexelor. Detaliile rafinate ale structurii de interacțiune ne pot scăpa la prima vedere. Forțându-ne să explicităm natura interacțiunii, teoria jocurilor poate scoate la iveală subtilități sau devianțe dintre cele mai neașteptate.

Jocuri secvențiale

Mă voi întoarce acum la jocurile în care agenții iau *decizii secvențiale* (voi discuta astfel de jocuri în detaliu în capitolul următor) și voi începe cu un exemplu simplu care demonstrează puterea teoriei jocurilor de a clarifica structurile de interacțiune, care, altminteri, vor fi doar vag înțelese.

În Figura 19.5, se confruntă două armate la granița statelor lor. Generalul I fie se poate retrage, păstrând *statu-quoul* (3, 3), fie poate invada. Dacă invadează, Generalul II poate fie să lupte, cu rezultatul (1, 1), fie să cedeze teritoriul disputat cu rezultatul (4, 2). Înainte ca I să ia o decizie, II poate fi în măsură să-și comunice intenția de a lupta în cazul în care este atacat, sperând să îl determine pe I să aleagă (3, 3), și nu (1, 1). Însă *această amenințare nu este credibilă*. I știe că, odată ce invadează, va fi în interesul lui II să cedeze, și nu să lupte. Acest echilibru unic este (4, 2). Acest concept de echilibru nu este conceptul static al „celui mai bun răspuns", pe care l-am discutat până acum. Mai degrabă, este un concept dinamic care începe în stadiile ulterioare ale jocului și merge spre stadiile anterioare. (Termenul tehnic este „inducția inversă".). Mai întâi, ne putem întreba ce ar fi rațional pentru II să facă dacă I invadează. Răspunsul „Cedează" duce la rezultatul (4, 2). Alegerea lui I este, prin urmare, între un curs de acțiune care duce la (3, 3) și unul care duce la (4, 2). Fiind rațional, îl alege pe cel din urmă.

După cum a observat Tucidide în *Războiul Peloponesiac*, promisiunile trebuie să fie credibile pentru ca cealaltă parte să își bazeze comportamentul pe ele.

> Jurămintele făcute pentru a sprijini orice fel de reconciliere erau valabile doar pe moment întrucât erau făcute de ambele părți doar în absența unei alte surse de putere pentru a depăși un impas; dar oricine găsea opoziția cu garda jos într-un anumit moment și folosea primul prilej pentru un atac în forță se bucura de o răzbunare mai dulce din exploatarea bunei-credințe, mai degrabă decât din câștigul în luptă directă. (...) Întrucât cuvântul nimănui nu era temeinic și niciun jurământ nu era destul de teribil pentru a aduce reconcilierea, toți aceia care se găseau într-o poziție superioară, gândindu-se că nici nu puteau spera să dobândească siguranța, se asigurau că evită daunele și nu-și permiteau să aibă încredere în cineva.

Cu alte cuvinte, persoana căreia i se face promisiunea ar trebui să se întrebe dacă ar fi rațional pentru acela care o face să se țină de cuvânt. Îngăduind comunicarea în Jocul Încrederii (capitolul 15), de exemplu, cel de al doilea jucător poate încerca să îl determine pe primul să facă un transfer mare promițând returnarea unui transfer la fel de mare. Dacă nu există ceva cu care să fie avut la mână, promisiunea nu este credibilă. În *Despre democrație în America*, Tocqueville face câteva comentarii sarcastice

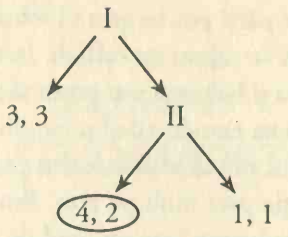

FIGURA 19.5

cu privire la o scrisoare a ministrului de război către indienii Cherokee, în care „susține că aceștia trebuie să renunțe la orice speranță de a rămâne în posesia teritoriilor pe care le ocupă în prezent, dar le oferă același asigurări pozitive odată ce au traversat Mississippi, de *parcă ar avea atunci puterea care acum îi lipsește*". Reforma economică în China a fost vulnerabilă la o problemă asemănătoare. Când guvernul a introdus reformele economiei de piață în agricultură, le-a promis fermierilor concesionări pe cincisprezece ani asupra terenurilor pentru a le da un stimulent să le îmbunătățească. Din moment ce un guvern autocrat nu poate fi tras la răspundere dacă nu își ține promisiunile, mulți fermieri nu au avut încredere și au folosit profiturile pentru consum. Un guvern autocrat *nu este capabil să se facă incapabil* să intervină.

Noțiunea credibilității este centrală în „cea de a doua generație" a teoriei jocurilor, care a început în jurul anului 1975. (Prima generație a început în jurul anului 1945.) Odată ce luăm ideea în serios, ni se cere să ne întrebăm cum ar putea agenții să *investească în credibilitate* pentru a spori eficiența amenințărilor și a promisiunilor. Există mai multe mecanisme. Unul presupune *consolidarea reputației*, de exemplu, investind în reputația de a fi într-o oarecare măsură sau din când în când irațional. Astfel, s-a spus că președintele Nixon, încurajat de Kissinger, a cultivat în mod deliberat un stil excentric pentru a-i face pe sovietici să creadă că ar fi putut acționa împotriva interesului american dacă l-ar fi provocat suficient. De asemenea, oamenii pot duce la îndeplinire amenințări când nu este în interesul lor să o facă pentru a-și consolida reputația de a fi duri, ceea ce îi va face credibili în fața celorlalți cu alte ocazii.

Un alt mecanism este cel de preangajare, discutat în capitolul 13. Acolo, preangajamentul a fost considerat o reacție rațională de ordin secund la înclinația agentului de a se comporta irațional. În context strategic, preangajamentul poate fi pe deplin rațional. În jocul descris în Figura 19.5, Generalul II poate construi o „mașinărie a apocalipsei", care să poată lansa automat un atac nuclear asupra țării cu care se află în conflict în cazul unei invazii. Dacă atât existența mașinăriei, cât și faptul că funcționarea acesteia nu stă în controlul țării II sunt fapte binecunoscute, atunci invazia ar putea fi împiedicată. Altminteri, II poate folosi strategia de „a da foc podurilor", anulând, cu alte cuvinte, orice posibilitate de retragere. Din nou, Generalul I ar fi descurajat dacă ar ști că Generalul II nu are altă soluție decât să lupte în caz de invazie.

În anumite cazuri, ambele părți pot încerca să folosească preangajamentul pentru a obține un avantaj competitiv în raport cu celălalt. În negocierile dintre muncitori și patronat, amenințările cu greva și boicotul s-ar putea să nu fie credibile. Patronatul știe că, din moment ce muncitorii au rate de plătit și familii de întreținut, nu își permit să facă grevă prea mult. Sindicatul știe că, din moment ce firma are contracte de onorat, nu își permite să țină producția prea mult pe zero. Pentru a spori credibilitatea amenințărilor, sindicaliștii ar putea investi într-un fond de grevă (poate împreună cu alte sindicate), iar patronatul ar putea investi în depozite mai mari. Sau negociatorii fiecărei părți ar putea să își enunțe pretențiile minime și ofertele maxime în public, asigurându-se că reputația lor va fi serios afectată dacă vor ceda. Un astfel de „joc al preangajării" ar putea fi o DP sau un joc al Laşului, în funcție de structura jocului ulterior.

◈◈◈

Notă bibliografică

Cea mai bună introducere elementară în teoria jocurilor este A. Dixit și S. Skeath, *Games of Strategy*, ediția a II-a (New York, Norton, 2004). Printre abordările avansate, sugerez cartea lui F. Vega-Redondo, *Economics and the Theory of Games* (Cambridge University Press, 2003). O trecere în revistă cu caracter enciclopedic și cu multe aplicații este R. Aumann și S. Hart, *Handbook of Game Theory with Economic Applications*, volumele 1-3 (Amsterdam, North-Holland, 1992, 1994, 2002). Aplicații pentru subiecte specifice se regăsesc în J.D. Morrow, *Game Theory for Political Scientists* (Princeton, NJ, Princeton University Press, 1994), și în D. Baird, H. Gertner și R. Picker, *Game Theory and the Law* (Cambridge, MA, Harvard University Press, 1994). Un studiu clasic al convențiilor îl reprezintă cartea lui D. Lewis, *Convention* (Cambridge, MA, Harvard University Press, 1969). Este inspirat, în mare parte, de o altă lucrare clasică, T. Schelling, *The Strategy of Conflict* (Cambridge, MA, Harvard University Press, 1960), în care a fost expusă pentru prima dată ideea punctelor focale. Lucrarea lui Schelling a pus bazele celei „de-a doua generații" în teoria jocurilor, dezvoltată formal de R. Selten, „Reexamination of the perfectness concept for equilibrium points in extensive games", *International Journal of Game Theory*, 4 (1975), pp. 25-55. Pentru diverse tehnici de preangajament în jocuri politice, vezi J. Fearon, „Domestic political audiences and the escalation of international disputes", *American Political Science Review*, 88 (1994), pp. 577-592. Pentru utilizarea lor în negocierea salariilor, vezi cartea mea *The Cement of Society* (Cambridge University Press, 1989).

Capitolul 20

JOCURILE ȘI COMPORTAMENTUL

Intenții și consecințe

Structura conceptuală a teoriei jocurilor este revelatoare. Dar oare ne ajută și să *explicăm comportamentul*? Să luăm un moment în considerare motivația pentru incendierea podurilor sau a navelor. Un astfel de comportament s-ar putea manifesta din motive strategice, cum este cazul în exemplul din ultimul capitol, dar și din alte motive. În *Oxford English Dictionary* există următorul citat din *Tarzan din neamul maimuțelor*, de E.R. Burroughs: „De teamă că ar putea să cedeze rugăminților acestui uriaș, ea a incendiat podurile lăsate în urmă." În acest caz nu avem de a face cu rațiuni strategice. Mai degrabă, femeia în cauză se teme că ar putea ceda rugăminților dacă nu face în așa fel încât cedarea să devină imposibilă. Chiar și în sfera militară, astfel de motivații care nu au în vedere un factor strategic ar putea fi la fel de importate ca motivațiile strategice. Un comandant poate incendia podurile pentru ca nu cumva frica de inamic să îi facă pe soldații săi să fugă de luptă. Acesta ar putea vrea chiar să împiedice *propria sa* dezertare, în cazul în care îi este teamă să nu cedeze slăbiciunii voinței. Comandantul A poate incendia podurile pentru a trimite un semnal comandantului inamic B că B nu poate conta pe faptul că trupele lui A vor fugi. Acesta se pare că a fost raționamentul lui Cortés când le-a spus marinarilor săi (în mod credibil, dar neadevărat) că flota sa nu poate face față unei călătorii pe mare, dând foc tuturor corăbiilor, mai puțin uneia. (De asemenea, incendiind corăbiile, putea să îi alăture pe marinari infanteriei sale.) Pentru a distinge între aceste explicații variate trebuie să determinăm intențiile agenților. Beneficiile actuale ale incendierii podurilor nu sunt nici necesare și nici suficiente pentru a pune pe picioare o explicație în termeni de beneficii așteptate (capitolul 3).

Cu toate că teoria jocurilor explică comportamentul prin apel la intențiile actorilor de a realiza anumite consecințe, ea poate să dea seamă de situații în care anumitor actori nu le pasă de consecințe. Ne putem gândi, de exemplu, la interacțiunea dintre Uniunea Europeană și noile țări membre din Europa de Est. Vechile state membre pot

fi tentate să impună condiții de aderare care să presupună subvenții în agricultură întotdeauna mai mici pentru noile state decât pentru statele vechi de dimensiuni comparabile. În termeni materiali, datorită beneficiilor obținute de pe urma aderării, noile state oricum o duc mai bine ca state membre de rang secund decât ca non-membre, cu toate că ar fi dus-o și mai bine dacă ar fi fost state cu drepturi depline. În termeni psihologici, insulta de a fi tratate ca inferioare le-ar putea determina să respingă astfel de condiții.[1] Anticipând o astfel de reacție, statele mai vechi ar putea fi determinate să ofere aderarea în termeni de egalitate deplină. Credința că statelor noi nu le pasă doar de considerente de ordin material le-ar putea asigura condiții materiale chiar mai bune.

Din moment ce nu am fost de față la negocierile de aderare, aceste observații sunt pur conjecturale. Știm, însă, că argumente de acest tip s-au oferit în cadrul Convenției Federale din Philadelphia, din 1787, în dezbaterea cu privire la termenii de aderare a viitoarelor state din Vest. Guvernatorul Morris și alții au propus ca aceste state să fie admise doar ca state de rang secund pentru a nu reuși vreodată să primească mai multe voturi decât cele treisprezece state inițiale. Împotriva acestei abordări, George Mason a argumentat puternic în favoarea admiterii noilor state cu aceleași drepturi precum cele vechi. În primul rând, a oferit un argument bazat pe principii: prin admiterea statelor din Vest în aceeași termeni legiuitorii ar face „ceea ce știm că este bine în sine". Acelora care ar fi putut să nu accepte argumentul le-a spus că ar fi puțin probabil ca noile state să accepte o propunere atât de umilitoare.

> Dacă ar fi să acceptăm statele din Vest în Uniune pe măsură ce acestea se formează, ele ar trebui tratate ca egale și nu ar trebui supuse unor discriminări umilitoare. Acestea vor avea aceeași mândrie și alte pasiuni pe care le împărtășim și fie nu se vor uni, fie vor părăsi repede o Uniune în care nu se află pe picior de egalitate din toate punctele de vedere cu confrații lor.

Mason se referă la „mândria și pasiunile" noilor state, și nu la interesul lor personal. Chiar dacă ar fi, de fapt, în interesul lor să adere la Uniune în condiții de inegalitate decât să rămână pe dinafară, acestea tot ar prefera, din ranchiună, să nu fie membre. În același timp, Mason face apel la interesul propriu al vechilor state, și nu la simțul lor pentru dreptate. În terminologia capitolului 4, acesta le spună că, întrucât noile state ar putea fi motivate de *pasiune mai degrabă decât de interes*, ar fi în

[1] În 2003, președintele Chirac a manifestat o astfel de atitudine când a răspuns sprijinului pe care l-au acordat politicienii din Europa de Est politicii americane din Irak, spunând că aceștia au ratat o mare ocazie de a-și ține gura, adăugând că aceștia nu erau, în mod evident, foarte bine-crescuți.

interesul vechilor state să acționeze ca și cum ar fi motivate de *rațiune mai degrabă decât de interes.*

Această situație a fost studiată în experimente prin intermediul Jocului Ultimatumului și al Jocului Dictatorului (Figura 20.1). În Jocul Ultimatumului, o persoană (Propunătorul) poate propune o distribuție (x, 10–x) a zece dolari între el și altă persoană (Respondentul). Ofertele nu pot fi făcute decât în dolari întregi. Dacă Respondentul acceptă, acea distribuție este pusă în practică. Dacă Respondentul respinge oferta, nimeni nu primește nimic. Cu toate că jocul a fost studiat în multe variante, mă voi concentra asupra interacțiunilor unice în condiții de anonimat. Întrucât subiecții interacționează prin terminale de calculator, aceștia nu cunosc identitatea partenerului lor. De multe ori li se aduce la cunoștință faptul că cel care coordonează experimentul nu va fi în măsură să determine cine face alegerile și în ce constau acestea, eliminând astfel posibilitatea ca deciziile lor să fie influențate de dorința de a-i face coordonatorului pe plac. Când subiecții joacă jocul de mai multe ori, niciodată nu se întâlnesc cu același partener, ceea ce îngăduie învățarea, dar nu și consolidarea reputației. În aceste condiții, scopul maxim al deciziilor este de a reflecta interesul personal fără nicio constrângere.

Presupunând că ambii agenți sunt raționali, concentrați pe interesul personal și au toate informațiile necesare cu privire la structura recompensei, iar aceste lucruri sunt binecunoscute, Propunătorul va oferi (9, 1), ofertă pe care Respondentul o va accepta. Dacă ofertele ar putea fi făcute în cenți, Propunătorul ar oferi (9,99, 0,01), care încă ar fi acceptată, întrucât ceva tot e mai bun decât nimic. În experimente, propunerile sunt de obicei în jur de (6, 4). Respondenții resping de regulă propunerile oferind 2

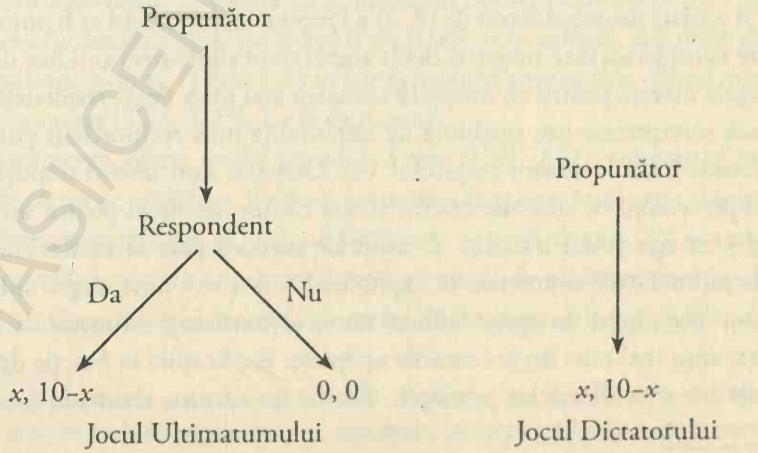

FIGURA 20.1

sau mai puțin.[1] Aceștia sunt dispuși să își dea singuri în cap. Evident, una dintre presupoziții este încălcată. În virtutea a ceea ce experimentul inițiază, putem exclude lipsa de informare sau lipsa unor lucruri binecunoscute cu privire la informație. Nu putem, însă, exclude eșecurile raționalității sau motivațiile care nu sunt generate de interesul personal.

Propunătorul ar putea fi un *altruist* care preferă o alocare întrucâtva egală în locul uneia în care el ia tot. Cu toate că altruismul pentru persoanele necunoscute care nu au vreo nevoie evidentă este o idee oarecum ciudată, este cel puțin coerentă cu raționalitatea. Putem respinge această ipoteză, însă, făcând o comparație între Jocul Ultimatumului și comportamentul în Jocul Dictatorului. În cel din urmă, care nu este în realitate un „joc", Propunătorul alocă unilateral o sumă de bani Respondentului, nelăsându-i celui din urmă vreo ocazie să răspundă. Dacă propunerile din Jocul Ultimatumului sunt dictate doar de altruism, alocările din Jocul Dictatorului ar trebui să fie dictate tot de altruism. În experimente, însă, comportamentul Propunătorilor este mult mai puțin generos în Jocul Dictatorului. Evident, comportamentul Propunătorului în Jocul Ultimatumului este determinat, cel puțin în parte, de așteptarea ca ofertele negeneroase să fie respinse.

Pentru a explica această respingere, am putea presupune că Respondenții vor fi motivați de *invidie* pentru a respinge ofertele sărace, iar Propunătorii egoiști, anticipând acest efect, vor face oferte suficient de generoase încât să fie acceptate. Dacă o astfel de explicație ar fi corectă, ar trebui să ne așteptăm ca frecvența respingerii la (8, 2) să fie cam aceeași ca atunci când Propunătorul este liber să propună orice alocare și când este constrâns – și știe că este constrâns – să aleagă între (8, 2) și (2, 8). În cadrul unor experimente, rata respingerii a fost mai mică în cel din urmă caz. Acest rezultat sugerează despre comportamentul Respondenților că este determinat de chestiuni de *echitate*. Și aceasta deoarece oferta de (8, 2) a Propunătorului, când ar fi putut să ofere (5, 5), este considerată mai incorectă decât atunci când alternativa ar fi fost dezavantajoasă în egală măsură pentru el. *Intențiile* contează mai mult decât rezultatele.

Această interpretare este susținută de importanța unei reciprocități puternice în jocuri precum Jocul Încrederii (capitolul 13). Oamenii sunt uneori dispuși să îi pedepsească pe ceilalți, cu anumite costuri și fără niciun beneficiu pentru sine, pentru că ceilalți s-au comportat incorect. O astfel de practică pare să încalce unul dintre canoanele raționalității enumerate în capitolul 12: în cazul unei alegeri între a acționa și a nu face nimic, un agent rațional nu va acționa dacă estimează că beneficiile așteptate sunt mai mici decât costurile așteptate. Explicațiile în funcție de altruism sau invidie nu ar încălca acest principiu. Pentru un altruist, rezultatul poate fi mai

[1] Spun „de regulă" din cauza unei variații semnificative la nivelul descoperirilor. Unele variații sunt influențate de sex, altele de cultură.

bun când aduce un beneficiu altcuiva cu un anumit cost pentru sine, iar pentru persoana invidioasă – când face un rău altcuiva cu un anumit cost pentru sine. Astfel de comportamente încalcă presupoziția interesului personal, dar nu și presupoziția raționalității. Dimpotrivă, explicația echității pare să încalce presupoziția raționalității. Reciprocitatea puternică determină un comportament asemănător cu acela pe care îl adoptăm când ne împiedicăm de un bolovan și îl lovim pentru a ne răzbuna: nu ne e de niciun folos, iar durerea crește.[1]

Inducția inversă

În Jocul Ultimatumului, jocul din Figura 19.5 și alte jocuri secvențiale, echilibrul este atins prin inducție inversă. În Jocul Ultimatumului, Propunătorul anticipează modul în care Respondentul ar putea reacționa la o propunere și își ajustează comportamentul în consecință. În aceste exemple, calculele implicate sunt destul de simple. În alte experimente, li se poate cere subiecților să realizeze lanțuri mai lungi de raționamente. De exemplu, doi subiecți pot trece prin trei runde de oferte și contraoferte pentru a împărți o sumă de bani, fiecare rundă presupunând o micșorare cu 50%.[2] În fiecare etapă, un agent poate fie să accepte suma propusă și să meargă în „dreapta", fie să facă o contraofertă și să meargă în „jos". Raționalitatea, interesul personal și lucrurile binecunoscute vor determina apoi următorul tip de raționament.

Persoana care face prima ofertă (Jucătorul I) va trebui să aibă în vedere dacă Jucătorul II preferă acea împărțire uneia în care ar primi o bucată mai mare dintr-o plăcintă mai mică. În același timp, Jucătorul I știe că Jucătorul II nu va face o propunere care îl va pune pe I într-o situație mai rea decât aceea în care s-ar afla dacă ar merge până în ultima rundă. În Figura 20.2, Jucătorul I poate primi cel puțin 1,25, luând tot ce a rămas în runda a treia. Jucătorul II nu poate, prin urmare, să îi ofere mai puțin de 1,25 în runda a doua, lăsând 1,25 ca sumă maximă pentru sine. Știind acest lucru, Jucătorul I va oferi (3,75, 1,25), iar II va accepta.

În experimente, oferta medie făcută de I este (2,89, 2,11), substanțial mai generoasă decât oferta de echilibru. Evident, avem de-a face cu o încălcare a uneia sau mai multora dintre presupoziții. (1) Primul jucător ar putea fi altruist. (2) Acesta s-ar putea teme că celălalt jucător ar putea să respingă oferta de echilibru pentru că nu este capabil să urmărească logica inducției inverse. (3) Ar putea el însuși să fie incapabil să

---

[1] O posibilă obiecție la această analiză este că, după cum am văzut în capitolul 15, oamenii pot obține plăcere din pedepsirea altora și, mai mult, așteptarea plăcerii poate fi ceea ce *motivează* pedeapsa.

[2] Micșorarea poate fi văzută ca un efect al deprecierii timpului (capitolul 6).

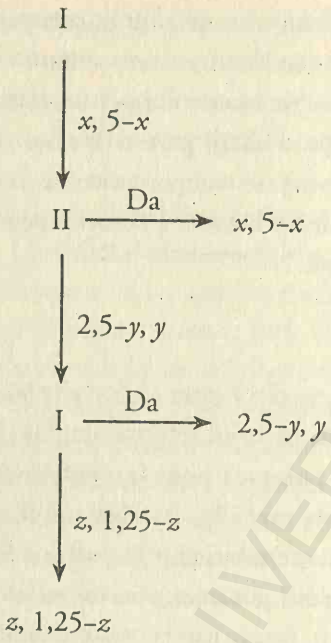

FIGURA 20.2

urmeze acea logică.[1] Prima și a doua ipoteză pot fi eliminate spunându-le subiecților că joacă împotriva unui calculator care este programat să răspundă optim. În acest caz, prima ofertă medie este (3,16, 1,84), care este substanțial mai generoasă decât oferta de echilibru. Din moment ce este greu de crezut că subiecții care fac oferte generoase au sentimente altruiste față de un calculator sau cred că este incompetent, trebuie să fie ei înșiși incompetenți.

Nu e ca și cum sarcina în sine ar fi una dificilă. Odată ce subiecților li se explică logica inducției inverse, aceștia se comportă impecabil în jocurile ulterioare. Mai degrabă, experimentul pune în evidență faptul că acest tip de raționament nu este natural pentru ființele umane. Chiar și un simplu raționament prematur poate fi ceva ce nu se întâmplă spontan, după cum pune în evidență Cursa Câștigătorului (capitolul 12). Sindromul „fratelui mai mic" (capitolul 18) are aceeași alură. Nu e ca și cum oamenii ar fi incapabili să înțeleagă în urma reflecției că ceilalți sunt raționali și capabili să delibereze la fel de mult ca ei, ci doar că tendința lor spontană este de a-i vedea pe ceilalți fixați mai degrabă în anumite obiceiuri, fără a fi capabili să se adapteze la mediu.

[1] Spre deosebire de ce se întâmplă în Jocul Ultimatumului, Jucătorul I nu are vreun motiv să se teamă că Jucătorul II ar putea să respingă oferta de echilibru ca incorect de mică, din moment ce Jucătorul I are și primul, și ultimul cuvânt.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

Anumite eșecuri ale teoriei jocurilor bazate pe alegerea rațională

Printre multe alte descoperiri care pun în evidență eșecurile predictive ale teoriei jocurilor, voi discuta „Dilema Prizonierului repetată finit", „Paradoxul Lanțului de Magazine", „Jocul Miriapodului", „Dilema Călătorului" și „Concursul de Frumusețe".

Când subiecții joacă mai multe DP succesive unul împotriva celuilalt și știu care este ultima rundă, se observă o proporție substanțială de alegeri de tip C, de multe ori peste 30%. O explicație intuitivă este aceea că un jucător poate alege C într-o rundă în speranța că și celălalt va face la fel („serviciu contra serviciu"). Însă, dacă jucătorii adoptă inducția inversă, aceștia vor înțelege că, în jocul final, ambii vor alege D întrucât nu va mai exista vreo ocazie pentru a influența comportamentul într-un joc ulterior. În penultimul joc, de asemenea, jucătorii vor alege D din moment ce comportamentul în jocul final este dat de argumentul anterior. Acest argument merge până la prima rundă, determinând trădarea în toate jocurile.

Un lanț de magazine este extins în douăzeci de orașe și în fiecare oraș se confruntă cu un potențial competitor. Competitorul trebuie să se decidă fie să înceapă o afacere împărțind piața cu lanțul de magazine, fie să stea în afara orașului. Lanțul de magazine are opțiunea fie de a răspunde în mod agresiv printr-o politică de reduceri de prețuri în vederea eliminării concurenței, ducând rivalul la faliment, dar suportând și anumite costuri, fie de a împărți piața. Recompensele sunt cele din Figura 20.3; primul număr în fiecare pereche este recompensa pentru potențialul concurent intrat pe piață.

Inducția inversă într-un singur joc produce (5, 5) ca în rezultatul de echilibru: rivalul intră pe piață și lanțul de magazine acceptă să împartă piața. Însă, gândindu-se la provocările de mai târziu, lanțul de magazine poate decide să se comporte mai agresiv și să își ducă rivalul la ruină, cu un anumit cost pentru sine, pentru a-i descuraja pe potențialii noi concurenți din alte orașe. Însă, dacă aplicăm inducția inversă șirului de douăzeci de jocuri, această strategie nu este viabilă. În al douăzecilea joc nu există vreun beneficiu de pe urma comportamentului agresiv, prin urmare, firma poate foarte bine

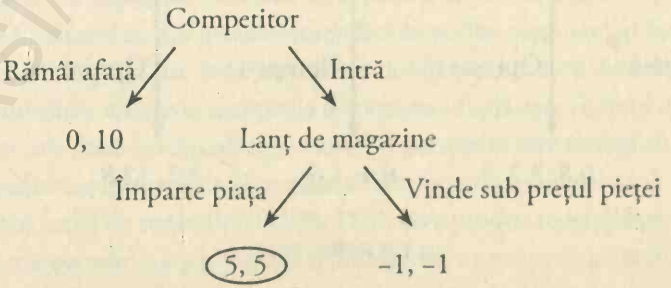

FIGURA 20.3

să împartă piața cu noul concurent. Acest lucru presupune, însă, că nu există beneficii de pe urma politicii de prețuri care vizează eliminarea concurenței nici în al nouăsprezecelea joc și așa mai departe, până la primul joc. Cu toate că dimensiunea acestui tip de politică pe pieţele reale este controversată, se manifestă uneori pe anumite piețe experimentale.

Jocul Miriapodului[1] este ilustrat în Figura 20.4 (recompensele exprimate în dolari). Inducția inversă îi spune Jucătorului I să aleagă „Oprește-te" la început, fiecare dintre cei doi rămânând cu o șaisprezecime din recompensa pe care ar fi putut să o obțină dacă ar fi continuat până la final. Într-un experiment obișnuit, 22% aleg „Oprește-te" la primul „nod" de alegere, 41% dintre cei care rămân aleg să se oprească la al doilea nod, 74% dintre cei care mai rămân aleg „Oprește-te" la al treilea nod, iar dintre cei care mai rămân, jumătate aleg să se oprească la al patrulea nod, iar jumătate aleg „Acționează". Devierea de la echilibrul (încercuit) prezis de inducția inversă este mare, cum este și creșterea medie a câștigurilor pentru jucători.

Pentru a explica aceste instanțe de cooperare și relație pradă-prădător aparent iraționale, am putea stipula existența unei incertitudini cu privire la anumite aspecte ale jocurilor. Jucătorii din viața reală se confruntă arareori cu un număr finit și cunoscut de runde. De cele mai multe ori, aceștia ar putea să creadă că interacțiunea va continua pentru o durată indefinită, astfel încât nu există vreo rundă finală de la care să se propage inducția inversă. În astfel de cazuri, adoptarea reciprocă a serviciului contra serviciu ar putea fi un echilibru în DP repetată. (Nu este un echilibru unic, întrucât și „Trădează întotdeauna, trădează întotdeauna" este, de asemenea, un echilibru. Din punct de vedere structural, avem de-a face cu ceva asemănător Jocului Asigurării, cu un echilibru bun și unul rău.) Dacă viața de zi cu zi determină un comportament de tipul serviciu contra serviciu, agenții îl pot aplica unor situații de laborator, în care nu este optim.

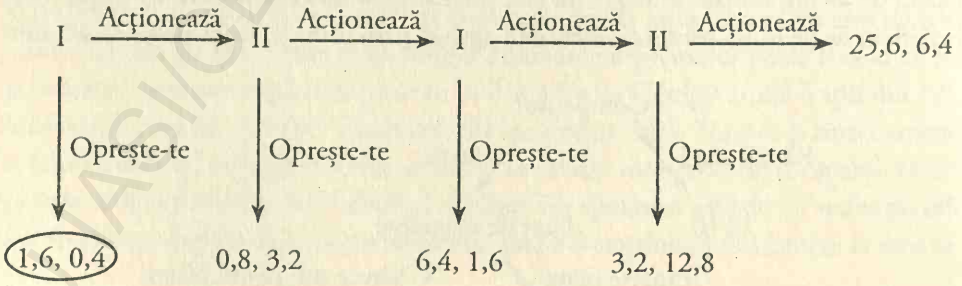

FIGURA 20.4

[1] Denumirea se referă la o versiune a jocului cu 100 de noduri.

În schimb, un agent poate fi nesigur cu privire la tipul de jucător cu care se confruntă. Să presupunem că este bine cunoscut faptul că, în societate, există și jucători irațional. Se știe că unii agenți vor coopera întotdeauna, că alții vor folosi serviciul contra serviciu în DP repetate finit, că alții vor folosi o politică de prețuri care vizează eliminarea concurenței pentru a-i descuraja pe nou-veniți chiar și în al douăzecilea oraș și așa mai departe. Nu se știe, însă, exact cine sunt acești indivizi. Există o probabilitate mai mare decât zero ca orice agent să fie irațional. În Paradoxul Lanțului de Magazine, un nou concurent potențial va fi descurajat dacă probabilitatea pe care o atribuie iraționalității directorului lanțului de magazine este suficient de mare. Directorul, cunoscând acest fapt, este stimulat să practice o politică de prețuri de eliminare a concurenței în raport cu primul nou-venit pentru a-i face pe ceilalți să creadă că este irațional. Când nou-veniții potențiali din alte orașe vor observa acest comportament, vor folosi un raționament bayesian (capitolul 11) pentru a atribui o probabilitate mai mare iraționalității acestuia. S-ar putea să nu fie suficient de mare pentru a-i descuraja, dar, dacă manifestă acest comportament în mod repetat, se va ajunge în cele din urmă la un nivel la care este mult mai rațional pentru ei să rămână afară. Un argument asemănător ar putea explica cooperarea în cazul unei DP repetate finit sau în cazul Jocului Miriapodului.

O altă posibilitate este ca, într-o Dilemă a Prizonierului repetată și în Jocul Miriapodului, cooperarea să posede ceva de felul unei calități a unui punct focal. Cu toate că indivizii *raționali* ar trăda cu prima ocazie, persoanele *rezonabile* nu ar face-o. Cu toate că această sugestie (asupra căreia voi reveni mai încolo) este destul de vagă, mi se pare mai în ton cu viața reală decât argumentele bazate pe incertitudinea cu privire la ce tip de persoană este celălalt. În primul rând, aceste argumente impun condiția ca jucătorii să întreprindă calcule extrem de complicate, care ocupă pagini întregi în manuale. În al doilea rând, introspecția și observația cauzală sugerează că, atunci când luăm decizii în viața de zi cu zi, nu ne gândim la ceilalți în acest mod. Am încredere în cineva cu o sumă mică de bani, dar nu cu una mai mare, și aceasta nu pentru că atribui o probabilitate mică caracterului său intrinsec de a fi de încredere, ci pentru că mă gândesc că pot avea încredere în respectiva persoană doar dacă mizele nu sunt foarte mari.

În Dilema Călătorului, doi jucători revendică în același timp același bagaj pierdut, făcând o cerere de despăgubire între optzeci de dolari și două sute de dolari. Pentru a descuraja revendicările excesive, compania aeronautică îi plătește fiecărui călător suma minimă pentru cele două cereri, adaugă o sumă R persoanei care revendică cel mai puțin, iar apoi scade aceeași sumă de la persoana care revendică cel mai mult. Să luăm în considerare două astfel de revendicări (100, 150), care produc recompensele (100 + R, 100 – R). Această pereche nu poate fi un echilibru din moment ce primul jucător ar fi stimulat să ceară 149, ceea ce ar duce la o recompensă de 149 + R, la care al doilea jucător ar răspunde cerând 148 și așa mai departe. După cum sugerează acest exemplu,

unicul echilibru se produce când ambii revendică 80. În experimente, acest rezultat este observat, de fapt, când R este mare. Când R este mic, însă, subiecții fac revendicări aproape de limita superioară de 200. Din nou, intuiția mea este că funcționează ceva de felul unui raționament de tip punct focal. Fiecare călător știe că, date fiind câștigurile rezultate în urma coordonării pentru o revendicare mai mare, ar fi o prostie să adopte strategia de echilibru și se așteaptă ca și celălalt jucător să știe acest lucru.

John Maynard Keynes a comparat bursa cu un Concurs de Frumusețe. El avea în minte concursurile populare de la vremea respectivă din Anglia, în care un ziar publica 100 de fotografii, iar oamenii votaü pentru șase poze care le plăceau cel mai mult. Cei care alegeau cea mai populară poză intrau automat într-o tombolă, în care puteau câștiga un premiu. Keynes scrie: „Nu avem de-a face cu un caz în care se aleg pozele care sunt, conform celei mai bune judecăți a celui care votează, cele mai frumoase și nici măcar cu cele care sunt, conform unei opinii comune, realmente cele mai frumoase. Am atins un al treilea nivel, în care ne folosim inteligența pentru a anticipa așteptările opiniei publice cu privire la opinia comună. Și cred că sunt unii care practică nivelul patru sau cinci sau care merg, poate, chiar mai sus de atât."

Într-un joc inspirat de observațiile lui Keynes, subiecții sunt rugați să aleagă un număr între 0 și 100. Jucătorul al cărui număr este cel mai aproape de două treimi din media numerelor alese câștigă un premiu dinainte stabilit. Media este constrânsă să fie 100 sau mai puțin, ceea ce implică faptul că două treimi din medie echivalează, prin constrângere, cu 67 sau mai puțin. Prin urmare, pentru orice medie care rezultă din alegerile celorlalți jucători, 67 va fi mai aproape de două treimi din acea medie decât orice număr peste 67. Dar când numerele sunt constrânse să fie 67 sau mai puțin, 2/3 din medie este constrânsă să fie 44 sau mai puțin și așa mai departe, până se atinge echilibrul unic 0. În experimente, foarte puțini subiecți aleg 0; numărul mediu este undeva în jurul lui 35. Pentru ca cineva să aleagă acest număr trebuie să creadă că și ceilalți, în marea lor majoritate, vor alege numere mari (sindromul fratelui mai mic). Faptul că acest număr reprezintă aproximativ două treimi din media tuturor numerelor, 50, sugerează că subiectul tipic ar putea crede că alții aleg numerele la întâmplare, în timp ce el este liber să optimizeze. În schimb, subiectul tipic poate crede că ceilalți trec prin două runde de eliminare, lăsându-l liber să optimizeze adăugând o a treia rundă.

Am sugerat că oamenii nu reușesc să se conformeze predicțiilor teoriei jocurilor, poate pentru că sunt *mai puțin raționali* sau *prea raționali*. Sindromul fratelui mai mic reprezintă, cu siguranță, un eșec al raționalității, la fel cum este și inabilitatea de a executa o simplă inducție inversă. A fi rezonabil înseamnă a ieși din capcanele întinse de raționalitate: a te concentra pe faptul că ambii jucători pot câștiga ignorând logica celui mai bun răspuns. După cum am mai spus, ultima idee este oarecum înrudită cu noțiunea de punct focal, însă doar oarecum. Punctele focale sunt echilibre, în timp ce

cooperarea din Dilema Prizonierului repetată finit, o revendicare prea mare în Dilema Călătorului sau alegerea opțiunii „Acționează" în Jocul Miriapodului nu sunt. Aceste alegeri au în comun cu alegerile punctului focal o proprietate greu de definit și foarte dependentă de context, care presupune un caracter evident și rezonabil.

Acest argument poate părea mai degrabă asemănător cu gândirea magică (capitolul 7) decât cu raționamentul punctului focal. A ignora cântecul de sirenă al raționalității înseamnă să urmăm dispoziția formulată de John Donne în *Aniversarea*:

> Cine-i mai în siguranță decât noi? când nimeni nu ne poate
> Trăda, decât unul dintre noi doi.
> Să ne păzim de temerile adevărate sau false.

A ignora temerile *adevărate* pare irațional sau magic. (Același lucru poate fi spus și despre ignorarea perspectivelor adevărate de câștig.) Eu prefer să văd lucrurile altfel: asemenea comportamente reflectă un standard mai înalt decât simpla raționalitate.[1] Acestea sunt chestiuni dificile, iar cititorii sunt invitați să se hotărască singuri cu privire la ele. O parte dintre aceste chestiuni vor forma subiectul capitolului următor.

◈◈◈

Notă bibliografică

C. Camerer, *Behavioral Game Theory* (New York, Russell Sage, 2004) este sursa pentru cele mai multe exemple din acest capitol. O analiză utilă a condițiilor în care predicțiile teoriei jocurilor standard eșuează este J.K. Goeree și C.A. Holt, „Ten little treasures of game theory and ten intuitive contradictions", *American Economic Review*, 91 (2001), pp. 1402-1422. Ideea, aparent simplă, a inducției inverse scoate la lumină paradoxuri profunde, unele dintre ele expuse în Introducerea cărții mele *The Cement of Society* (Cambridge University Press, 1989). Jocul ilustrat în Figura 20.2 este din E. Johnson *et al.*, „Detecting failures of backward induction", *Journal of Economic Theory*, 104 (2002), pp. 16-47. Dilema Călătorului este luată din K. Basu, „The traveler's dilemma: Paradoxes of rationality in game theory", *American Economic Review: Papers and Proceedings*, 84 (1994), pp. 391-395.

---

[1] Nu mă refer la ideea de *rațiune* (capitolul 4) în sensul motivațiilor imparțiale, ci la o noțiune ceva mai obișnuită. Ideea (nedezvoltată) pe care o am în minte se leagă de faptul că oamenii cooperează mai des în Dilema Prizonierului când aceasta este descrisă ca un „Joc al comunității" decât atunci când este descrisă ca un „Joc de tip Wall Street".

Capitolul 21

ÎNCREDEREA

Slăbirea poziției de apărare

Egoismul, spune Tocqueville, este „rugina societății". Similar, se spune adesea că încrederea este „vaselina societății"[1]. Viața de zi cu zi ar fi insuportabil de dificilă dacă nu am avea încredere în ceilalți că vor face ceea ce au spus că vor face, cel puțin într-o anumită măsură. Cu toate că cercetătorii au definit încrederea în felurite moduri, eu voi folosi o definiție comportamentală simplă: a avea încredere în cineva înseamnă a slăbi poziția de apărare, *a nu lua măsuri de precauție împotriva unui partener de interacțiune*, chiar dacă celălalt, din cauza oportunismului sau incompetenței, ar putea să acționeze într-un mod care să justifice anumite precauții.[2] Înțeleg prin „oportunism" un interes personal miop sau „brut", neconstrâns nici de considerente etice, nici de considerente care țin de prevedere. Acțiuni oportuniste tipice care ar putea justifica precauțiile celorlalți sunt minciuna, copiatul la examen, încălcarea unei promisiuni, delapidarea, înșelarea partenerului de viață sau alegerea strategiei necooperative în Dilema Prizonierului.

Cineva poate să aibă sau să nu aibă încredere *în sine* că va respecta o negociere, că se va abține de la alcool sau că nu va abate corabia de la cursul său chiar dacă va auzi cântecul sirenelor. Neîncrederea în sine este evidențiată de preangajamente sau de construcția de reguli private (capitolul 13). Aceste strategii pot fi, însă, foarte costisitoare din cauza efectelor de semnalizare. Dacă ceilalți observă o instanță a unui astfel de comportament precaut față de sinele meu viitor, ei pot infera incorect, după cum am văzut în capitolul 10, că îmi lipsește autocontrolul în general. Prin urmare, pot ezita să aibă încredere în mine în cazul în care (1) lipsa mea de autocontrol poate fi

---

[1] În literatura de specialitate mai veche despre dezvoltarea economică, *corupția* era uneori considerată vaselina societății.

[2] Încrederea înțeleasă astfel presupune o *dublă abținere*, una dintre părți stăpânindu-și precauțiile în speranța că și cealaltă își va stăpâni comportamentul oportunist.

costisitoare pentru ei, (2) nu sunt disponibile instrumente de preangajare și (3) regulile private sunt irelevante, ca în cazul unor întâlniri unice. În multe societăți există norme privitoare la abstinența totală de la alcool, așa cum există și norme împotriva beției (capitolul 22).

Neîncrederea se poate manifesta în două moduri. Pe de o parte, cineva se poate abține pur și simplu să interacționeze cu un partener potențial dacă interacțiunea l-ar putea face vulnerabil la incompetență sau oportunism. Pe de altă parte, cineva se poate angaja în interacțiune asigurându-se că și-a luat toate precauțiile împotriva acestor riscuri. Prin urmare, încrederea este rezultatul a două decizii succesive: angajarea în interacțiune și abținerea de la monitorizarea partenerului de interacțiune. Întrucât decizia de a se abține de la interacțiune este greu de observat, se poate foarte ușor subestima nivelul de neîncredere la nivelul unei societăți. S-ar putea crede cu ușurință că există mai multă neîncredere într-o societate în care oamenii se urmăresc în permanență decât într-una în care sunt retrași în sine. Însă, la o analiză mai atentă, s-ar putea descoperi că ultima societate este foarte ineficientă în ceea ce privește tranzacțiile aducătoare de câștig reciproc, care nu sunt însă efectuate.

Montaigne a oferit o descriere optimistă: „Când călătoresc, oricine îmi are punga în păstrare este răspunzător de ea fără niciun fel de supraveghere." Alte instanțe ale încrederii pot presupune abținerea de la anumite acțiuni, cum ar fi:

- A citi jurnalul partenerului de viață
- Folosirea supraveghetorilor pentru monitorizarea studenților în timpul unui examen
- Verificarea scrisorilor de recomandare ale unui potențial angajat
- Impunerea unei garanții asupra unui chiriaș
- Insistența de a face contracte scrise cu titlu executoriu
- A-i cere partenerului mai puțin bogat să semneze un contract prenupțial
- Ascunderea banilor de copii
- Încuierea ușii când pleci de acasă
- Preangajarea de a-i pedepsi pe trădători într-o DP
- Consultarea unui al doilea medic sau a unui alt mecanic auto

După cum am observat deja, obiectul încrederii poate fi *abilitatea* celorlalți oameni sau *motivațiile* lor. Distincția este viu ilustrată în istoria mișcărilor de rezistență. În țările aflate sub ocupație nazistă în al Doilea Război Mondial, se întâmpla din când în când ca membrii mișcărilor de Rezistență să fie uciși deoarece se credea că erau agenți germani. Se întâmpla, deși mult mai rar, să fie uciși pentru că nu puteau fi crezuți că își vor ține gura. Ar putea ieși la iveală că cineva este bețiv și să fie executat de Rezistență pentru a nu da în vileag informații periculoase în stare de ebrietate. Pentru

a da un exemplu ceva mai obișnuit, mă pot îndoi de abilitățile unui mecanic auto sau îi pot pune onestitatea sub semnul întrebării. Când cer a doua opinie medicală, de mult ori o fac îngrijorat de competența medicului sau mă pot gândi că îmi recomandă o intervenție chirurgicală inutilă pentru a-și umple buzunarul. În cele ce urmează voi discuta în mare măsură chestiunea onestității.

Motive pentru încredere

Există mai multe motive pentru care oamenii se pot stăpâni să își ia precauții.[1] (1) Costul luării de precauții poate depăși beneficiile așteptate, fie cu o anumită ocazie, fie de-a lungul vieții ca întreg. Dacă în satul în care locuiesc lucrează un mecanic auto și ar trebui să călătoresc 80 de kilometri cu taxiul pentru o a doua opinie, s-ar putea ca acest lucru să nu merite efortul. În general, viața e prea scurtă pentru a te teme tot timpul că lumea ar putea să profite de tine. Pierderile ocazionale care survin din faptul de a avea încredere în cineva în care nu merită să ai încredere sunt mici prin comparație cu liniștea pe care o aduce lipsa oricărei griji. (2) Însăși acțiunea de a-ți lua măsuri de precauție poate furniza informații care pot fi exploatate de oportuniști. Montaigne citează maxima latină *Furem signata sollicitant. Aperta effractarius praeterit* („Casele încuiate invită hoțul: acesta trece pe lângă ele dacă le vede larg deschise"). (3) Ideea luării de precauții poate fi incompatibilă cu atitudinea emoțională a agentului față de cealaltă persoană. Când oamenii sunt îndrăgostiți, aceștia pot refuza să se angajeze în calculele seci pe care le implică un contract prenupțial. Strofa din Donne citată în capitolul 20 este binevenită și în acest context: „Să ne păzim de temerile adevărate sau false." (4) Pot avea credințe anterioare cu privire la gradul de încredere al celeilalte persoane. (5) Pot încerca să induc încrederea având încredere în persoana respectivă.

În cele ce urmează mă voi concentra asupra punctului (4) și, pe scurt, asupra punctului (5). În timp ce mulți cercetători definesc încrederea exclusiv în termenii lui (4), eu cred că faptul de a ne concentra pe reținerea deliberată are avantajul de a pune în lumină interacțiunea dintre acela care are încredere și acela în care se are încredere. Dacă persoana în care se are încredere *percepe* lipsa unor precauții, această percepție ar putea-o determina să acționeze altfel decât ar face-o altminteri. În cazul (2), acest lucru se întâmplă pentru că persoana presupune că nu va exista nicio *ocazie* pentru

[1] Voi folosi a se „stăpâni" (*refrain*) pentru a indica abținerea *deliberată*. În unele cazuri, pe care urmează să le discut, ideea luării de precauții, de pildă, citind jurnalul partenerului de viață, se poate să nu fi trecut niciodată prin mintea persoanei. Însă aceasta nu este „încrederea oarbă", pe care o voi defini mai târziu, dacă agentul a avut *oportunitatea* de a-și lua măsuri de precauție. Pentru cealaltă persoană implicată în relație, faptul că persoana a avut oportunitatea, însă nu s-a folosit de ea, este un semn care spune multe, indiferent dacă abținerea este percepută ca deliberată sau nu.

un comportament oportunist. În alte cazuri, pe care le voi aborda mai târziu, percepția poate modifica motivația cuiva de a se comporta oportunist, reflectând o intuiție preanalitică potrivit căreia încrederea are o anumită calitate de a deveni realitate în virtutea faptului că este așteptată. Același lucru este valabil și pentru neîncredere. După cum a observat Proust, „odată ce gelozia este descoperită, ea este considerată de obiectul ei ca fiind o lipsă de încredere care îi dă dreptul de a ne înșela".

Motive pentru a fi demn de încredere

Oamenii pot fi percepuți drept *demni de încredere* din mai multe motive. Voi discuta patru dintre acestea: comportamentul trecut, stimulentele, semnele și semnalele. De multe ori știm – sau credem că știm (vezi capitolul 10) – din observarea celorlalți oameni că aceștia își țin în mod constant promisiunile, că se abțin de la minciună, că au grijă de lucrurile care nu le aparțin și așa mai departe. Mai mult, o persoană care se știe demnă de (ne)încredere va tinde să îi considere și pe ceilalți drept demni de (ne)încredere. După cum spune La Bruyère, „ticăloșii îi cred cu ușurință pe ceilalți la fel de răi ca ei înșiși; aceștia nu pot fi păcăliți și nici ei nu păcălesc pentru mult timp."[1] Există dovezi venite pe cale experimentală că acest mecanism chiar funcționează. Invers, A poate avea încredere în C pentru că știe că B, în care are încredere, are încredere în C. Însă inferența poate să nu fie validă pentru că încrederea lui B în C se poate datora doar unui fals efect de consens. După cum ilustrează aceste exemple, de multe ori se întâmplă să avem încredere sau să nu avem încredere în oameni din motive greșite, crezând că ceilalți sunt fie mai mult ca noi, fie „mai mult ca ei înșiși" (mai consecvenți în comportamentul lor) decât sunt de fapt.

În comunitatea restrânsă a negustorilor de diamante, unde tentația pentru comportament oportunist este enormă, un acord verbal fără martori este la fel de constrângător ca un contract scris. Un negustor care încalcă un acord poate avea de câștigat pe termen scurt, dar pe urmă va fi evitat pentru totdeauna de toți ceilalți negustori.[2] Acesta nu ar mai putea nici să își lase moștenire afacerea copiilor săi, după cum se obișnuiește adesea în această comunitate. În cazul negustorilor din New York, care sunt în mare parte localizați în comunitățile ultraortodoxe evreiești, un escroc ar fi ostracizat social. Ultimul mecanism consolidează calitatea de a fi demn de încredere, dar

---

1 La Rochefoucauld are o opinie diferită: „Oamenii sunt înșelați mult mai ușor atunci când încearcă să-i înșele pe alții."

2 Acesta nu este un mecanism de tip serviciu contra serviciu, în care jucătorul care trădează într-o rundă și este pedepsit runda următoare poate fi iertat dacă reia comportamentul cooperant, ci este, mai degrabă, un mecanism de declanșare a „soluției crunte", în care o singură trădare împiedică iertarea, chiar dacă agentul se comportă bine ulterior.

nu este necesar pentru aceasta. Stimulentul de a păstra o anumită reputație de dragul onestității și a faptului de fi demn de încredere este de cele mai multe ori suficient.

Semnele sunt *trăsături* ale indivizilor despre care se crede, pe drept sau nu, că indică un caracter demn de încredere. Într-un studiu cu privire la ce anume îi face pe șoferii de taxi să aibă încredere că clienții lor nu îi vor jefui sau ataca, femeile sunt considerate mai demne de încredere decât bărbații, oamenii în vârstă mai demni de încredere decât tinerii, albii mai de încredere decât negrii, cei bogați mai demni de încredere decât cei săraci, cei închiși în sine mai demni de încredere decât cei vorbăreți și cei sinceri mai demni de încredere decât cei șmecheri. Un șofer spaniol de taxi din New York i-ar considera pe pasagerii spanioli mai demni de încredere decât pe aceia din alte grupuri etnice. Șoferii catolici din Belfast i-ar găsi mai demni de încredere pe catolici decât pe protestanți și invers, în cazul șoferilor de taxi protestanți. Trăsături mai generice sunt ochii care nu sunt prea apropiați unul de altul și privitul interlocutorului în ochi.

Semnalele sunt *comportamente* care dau dovezi ale calității de a fi demn de încredere. Acestea pot include producerea deliberată sau imitarea semnelor. De exemplu, se pare că o foarte bună modalitate de a genera o privire sinceră este să te concentrezi la baza nasului interlocutorului. În acest caz, semnalul va funcționa dacă persoana cealaltă crede că o privire sinceră este un indicator credibil al comportamentului și nu știe cât de ușor se poate truca. Alte comportamente funcționează ca semnale dacă sunt prea costisitoare pentru a fi accesibile indivizilor care nu sunt demni de încredere. Pentru a falsifica cu succes o semnătură e nevoie de mult exercițiu, pe când a realiza propria semnătură nu presupune niciun efort. Un om sărac se poate îmbrăca după modelul unui bancher de pe Wall Street pentru a părea demn de încredere în fața unui șofer de taxi, dar este puțin probabil că ar face așa ceva din moment ce costurile pe care le comportă ar fi mai mari decât ce se poate aștepta să câștige din tâlhărie. Dimpotrivă, a flutura *Wall Street Journal* pentru a chema un taxi este un lucru pe care și-l permite oricine, prin urmare, nu poate fi folosit pentru a distinge între aceia în care poți avea încredere și aceia în care nu poți avea încredere. În măsura în care încrederea se bazează pe credința că partenerul de interacțiune are un orizont lung de timp (o rată mică a deprecierii timpului), manifestările costisitoare ale condiției fizice, precum și suplețea pot servi ca semnal, dată fiind credința (falsă) că acuitatea, ca trăsătură de caracter, fie se aplică în toate cazurile, fie nu se aplică deloc.

De multe ori, oamenii au încredere în ceilalți pentru că îi consideră motivați de altceva decât de interesul personal. Alteori, însă, avem încredere în oameni doar dacă îi considerăm ca având un interes egoist. În *Șoimul maltez*, dl. Gutman îi spune lui Humphrey Bogart: „Nu am încredere într-un om care nu are grijă de el însuși." Napoleon a spus că Talleyrand nu putea fi de încredere pentru că nu cerea niciodată favoruri pentru familia sa. Se spune că François Mitterrand era la fel de neîncrezător în

aceia care nu îi cereau niciodată favoruri. În general, o mare problemă pentru șarlatani și pentru aceia care abuzează de încrederea oamenilor este să își convingă victimele de faptul că acționează din interes personal. Să presupunem că fac o plimbare alături de cineva și îi spun că poate face o avere investind o mică sumă de bani. Prima întrebare care mi s-ar pune ar fi: „Oare nu e prea frumos ca să fie adevărat?" A doua întrebare ar fi: „Dacă chiar este adevărat, de ce vrei să împarți această oportunitate cu mine în loc să ții totul pentru tine?" Escrocul experimentat poate să provoace încrederea victimei sale spunându-i o poveste plauzibilă cu privire la interesul personal care îl determină să împartă o parte din câștiguri. În absența unei istorii anterioare de interacțiune, afirmarea unor motivații binevoitoare nu este credibilă.

La fel cum oamenii pot fi (considerați) mai mult sau mai puțin demni de încredere, tot astfel aceștia pot fi mai mult sau mai puțin *încrezători*. Altfel spus, dacă atât A, cât și B au aceeași opinie despre C (sau nu au nicio opinie), A poate avea încredere în C, iar B poate să nu aibă. Predilecția pentru încrederea în ceilalți este mai cu seamă importantă în demararea unor întreprinderi cooperative. În cazul interacțiunilor repetate, cooperarea poate fi sprijinită de reciprocitate *cu excepția primei runde*, în care nu există o istorie prealabilă de interacțiune. Pentru a putea începe, părțile trebuie să coopereze necondiționat în prima rundă. Un individ încrezător va urma „serviciul contra serviciu": va coopera în prima rundă și va întoarce favorul în rundele ulterioare. După cum spune proverbul, „M-ai păcălit o dată, să-ți fie rușine; m-ai păcălit de două ori, să-mi fie rușine". O persoană neîncrezătoare va proceda invers: va trăda în prima rundă și va întoarce favorul în celelalte runde. În interacțiunile care presupun două persoane, este nevoie de doi indivizi încrezători pentru demararea cooperării.

Cum poate încrederea să determine un caracter demn de încredere

În cazul în care lumea are încredere în tine, acest lucru poate determina un caracter demn de încredere dacă ceilalți *știu* că te-ai stăpânit de la a-ți lua precauțiile pe care ți le-ai fi putut lua. Referindu-se la servitorul care se ocupa de punga sa, Montaigne scria: „Acesta mă poate înșela la fel de bine ca atunci când aș ține contabilitatea și doar dacă nu e vreun diavol, printr-o astfel de încredere lipsită de precauții îl fac să-mi fie dator cu onestitatea." Această formă de reciprocitate este diferită de aceea reprezentată de strategia „serviciu contra serviciu" din DP. „Serviciu contra serviciu" poate fi o strategie de echilibru când jocul este jucat de un număr indefinit de ori, iar cele două alegeri din orice joc sunt făcute simultan. Dimpotrivă, observația lui Montaigne se aplică în cazul unor jocuri unice, în care o parte face o anumită alegere și o face cunoscută înainte de alegerea celeilalte părți. Experimentele despre înclinația de a coopera în DP unice confirmă această intuiție. Într-un studiu, subiecților li s-a comunicat alegerea celuilalt jucător înainte ca ei să facă alegerea proprie. Când jucătorii au știut de

trădarea celeilalte părți, doar 3% au ales să coopereze. Când au știut de cooperarea celeilalte părți, 16% au cooperat.

Putem menționa un alt exemplu pentru a scoate în evidență caracterul rafinat al încrederii. Acest Joc al Încrederii este jucat cu două condiții. În ambele, „investitorul" are opțiunea de a transfera orice sumă între 0 și 10 din proprietatea sa de 10 unități monetare (UM) către un „administrator". Coordonatorul experimentului triplează apoi suma trimisă, astfel încât, dacă investitorul trimite 10, administratorul primește 30. Administratorul poate apoi să decidă să returneze investitorului orice sumă între 0 și suma mărită (30 în exemplul de față). În cele din urmă, dacă investitorul se decide să facă un transfer, trebuie să enunțe și suma pe care și-ar dori să o primească înapoi de la administrator.

Aceste trăsături definesc eticheta improprie (după cum voi argumenta) de „condiție de încredere". În cazul „condiției-stimulent", investitorului îi este oferită opțiunea de a enunța, la momentul la care face transferul și anunță suma care trebuie returnată, că va amenda administratorul cu 4 UM în cazul în care transferul returnat nu acoperă suma dorită. Unii investitori folosesc această opțiune, alții nu. Dacă nu o folosesc, administratorii știu că investitorul avea acea opțiune, dar s-a abținut de la folosirea ei. Descoperirile experimentale arată că suma cea mai mare returnată este determinată de condiția-stimulent în care nu se impune vreo amendă și este cea mai mică atunci când este impusă o amendă, returnările bazate pe încredere aflându-se la un nivel intermediar. Acest efect este anticipat de investitori, care investesc cam 30% mai mult pe condiția „stimulent, fără amendă" decât în oricare dintre celelalte două scenarii.

Condiția „stimulent, fără amendă" corespunde definiției pe care o dau eu încrederii. Ceea ce experimentatorii au numit „condiția încrederii" eu aș numi mai degrabă *încredere oarbă*. Aceasta se manifestă când precauțiile sunt *excluse* și este distinctă de situația în care precauțiile *nu sunt alese*. Descoperirea extraordinară este aceea că încrederea (care nu este oarbă) determină o cooperare mai mare decât încrederea oarbă. Slăbirea poziției de apărare este cea care face diferența.

Oamenii pot avea încredere sau neîncredere în *instituții*. Pot avea încredere că o bancă nu va da faliment sau preferă să își țină economiile la saltea. Pot avea încredere într-un tribunal care să fie imparțial într-o dispută între vecini sau pot prefera să își facă singuri dreptate. Strict vorbind, aceste cazuri nu se conformează definiției pe care am dat-o la început din moment ce nu există (sau există foarte puține) precauții pe care cetățenii să le ia atunci când au de-a face cu o instituție, cu excepția cazurilor în care refuză să aibă de-a face cu ea. Când am încredere în bancă, aceasta este o formă de încredere oarbă. Într-o anumită măsură, încrederea mea afectează șansele ca ea să nu dea faliment, însă numai pentru că, într-o oarecare măsură, contribuția mea la rezervele băncii o face mai puțin vulnerabilă. Dimpotrivă, încrederea lui Montaigne în

servitorul său determină în cel din urmă un caracter demn de încredere prin ruşinea sau vina pe care ar resimți-o dacă ar trăda acea încredere.

Însă, după cum însuși Montaigne a susținut, aceste consecințe benefice ale încrederii în ceilalți sunt esențialmente produse derivate (capitolul 4). Sunt șanse mici că vei avea succes dacă te stăpânești de la luarea unor precauții pentru unicul scop de a-i face pe ceilalți să se comporte frumos. „Este un excelent mod de a câștiga inima și mintea unui alt om acela de a avea încredere, punându-te la dispoziția sa, cu condiția ca acest lucru să se întâmple de bunăvoie, fără constrângeri impuse de necesitate și cu condiția ca încrederea să fie curată și pură, iar fruntea noastră să nu fie încrețită de ezitări."

◈◈◈

Notă bibliografică

În D. Glaeser *et al.*, „Measuring trust", *Quarterly Journal of Economics*, 115 (2000), pp. 811-846, sunt oferite dovezi că oamenii demni de încredere sunt chiar aceia care au încredere în alții. Comunitatea negustorilor de diamante din New York este analizată de B. Richman, „Community enforcement of informal contracts: Jewish diamond merchants in New York" (prelegere, Harvard University, Olin Center for Law and Economics, 2002). Utilizarea semnelor și semnalelor de către șoferii de taxi pentru a determina caracterul demn de încredere al pasagerilor este abordată de D. Gambetta și H. Hamil, *Streetwise* (New York, Russell Sage, 2005). Modurile prin care se poate conferi credibilitate unei înșelătorii sunt analizate într-o carte neglijată a lui N. Leff, *Swindling and Selling* (New York, Free Press, 1976). Experimentul cu privire la cooperare în Dileme ale Prizonierului unice este relatat în E. Shafir și A. Tversky, „Thinking through uncertainty: Nonconsequentialist reasoning and choice", *Cognitive Psychology*, 24 (1992), pp. 449-474. Rezultatele Jocului Încrederii descrise în text se regăsesc în E. Fehr și B. Rockenbach, „Detrimental effects of sanctions on human altruism", *Nature*, 422 (2003), pp. 137-140.

BCU IASI - UNIVERSITY LIBRARY

Capitolul 22

NORMELE SOCIALE

Conștiința colectivă

Sociologii se referă uneori la „conștiința colectivă" a unei comunități ca la setul de valori și credințe împărtășit (și despre care se știe sau se crede că este împărtășit) de membrii săi. Din punctul de vedere al valorilor, conștiința colectivă include normele morale și sociale, religia și ideologiile politice. Din punctul de vedere al credințelor, include opinii cu privire la chestiuni de fapt, precum și cu privire la relațiile cauzale dintre fapte, de la zvonuri despre comerțul cu sclavi albi până la efectele perverse ale beneficiilor aduse de șomaj. În acest capitol mă voi ocupa de normele sociale și de modul lor de funcționare. În capitolul următor mă voi ocupa de formarea colectivă sau, mai bine zis, interactivă a credințelor. Există o dublă asimetrie în modul în care abordez valorile și credințele. Pe de o parte, nu am prea multe de zis despre apariția normelor sociale, nu pentru că problema nu este interesantă, ci pentru că este prea grea. Pe de altă parte, nu am prea multe de spus nici despre substanța credințelor populare sau colective. Conținutul lor poate varia destul de mult în timp și spațiu, pe când mecanismele de apariție, propagare, schimbare și distrugere a credințelor variază mult mai puțin.

Funcționarea normelor sociale

Să luăm în considerare două enunțuri:

> Poartă întotdeauna haine negre dacă soarele este puternic.
> Poartă întotdeauna haine negre la o înmormântare.

Prima dispoziție este o chestiune care ține de raționalitatea instrumentală întrucât aerul dintre corp și haine circulă mai rapid în cazul hainelor negre. A doua exprimă o normă socială care nu are vreo semnificație instrumentală evidentă. Existența și importanța normelor sociale nu poate fi pusă la îndoială. Cauzele proxime implicate

BCU IASI CENTRAL UNIVERSITY LIBRARY

în funcționarea lor sunt cunoscute destul de bine, însă originea lor ultimă, precum și funcția lor ultimă sunt chestiuni controversate.

O normă socială este o dispoziție de a acționa sau de a te reține de la vreo acțiune. Unele norme sunt necondiționate: „Fă X, nu face Y."[1] Printre aceste norme sunt incluse aceea de a nu consuma carne de om, de a nu avea relații sexuale cu fratele sau sora, de a nu trece în fața altcuiva la coadă, de a nu purta haine roșii (după cum le sfătuiesc unele mame pe fiicele lor), de a purta negru la înmormântare, de a folosi tacâmurile la masă într-un anumit mod, de a trata mai întâi pacientul mai grav bolnav. Alte norme sunt condiționate: „Dacă faci X, atunci fă Y" sau „În cazul în care ceilalți fac X, atunci fă X". În multe grupuri există o normă care spune că persoana care a sugerat ce acțiune să fie întreprinsă are responsabilitatea executării ei[2]; prin urmare, multe sugestii valoroase sunt astfel pierdute. Un cuplu fără copii poate fi supus unei norme conform căreia prima persoană care sugerează că ar trebui să facă un copil va avea o responsabilitate mai mare pentru creșterea acestuia; ca rezultat, multe cupluri care ar dori să aibă un copil pot rămâne fără copii.[3] Poate că nu există vreo normă care să îmi spună să trimit felicitări de Crăciun verilor mei, dar, odată ce am început să fac acest lucru, există o normă care îmi spune să continui și care le spune verilor mei să întoarcă gestul. Cu toate că sunt condiționate, aceste norme nu sunt condiționate de vreun *rezultat* care urmează a fi realizat prin acțiune, cum este dispoziția de a purta negru dacă soarele este puternic.

Voi da exemple și mai încolo. Însă, în primul rând, trebuie să spun ceva despre ce anume conferă eficiență cazuală normelor sociale și prin ce sunt acestea diferite de alte norme. Un răspuns la îndemână pentru prima întrebare este că normele sociale funcționează prin intermediul unor sancțiuni informale direcționate înspre aceia care încalcă normele. În mod obișnuit, sancțiunile afectează situația materială a contravenientului fie printr-un mecanism al pedepsei directe, fie prin pierderea unor oportunități cauzate de ostracizarea sa socială. Un fermier care încalcă normele comunității se poate trezi cu hambarul incendiat sau cu oile sacrificate. Sau poate descoperi că vecinul său refuză să îl ajute cu recolta. Mecanismul *bârfei* poate funcționa ca un multiplicator asupra acestor sancțiuni, adăugând sancțiuni ale terților la pedepsele inițiale.

Putem lua în considerare ce ar face un crescător de vite dacă animalul unui vecin i-ar intra de mai multe ori pe proprietate. Acesta ar putea să sechestreze animalul, asigurându-și astfel un beneficiu pentru sine și o pierdere pentru vecin. Ar putea aduce anumite daune animalului pentru a-i reduce valoarea (de exemplu, un taur ar putea

---

[1] În cele ce urmează, „condiționat" și „necondiționat" se referă la conținutul normei sociale. După cum am remarcat în capitolul 5, toate normele sociale sunt condiționate în sensul că funcționarea lor este condiționată de prezența unui observator.

[2] Această normă ar putea fi legată de raționamentul punctului focal.

[3] Prin urmare, norma determină un Joc al Lașului.

fi castrat), fără a beneficia direct de acest lucru, dar prejudiciindu-și vecinul. Ar putea duce animalul într-un loc îndepărtat, determinând astfel un anumit cost și pentru sine, și pentru vecin. Sau ar putea să întrerupă toate relațiile cu vecinul său (ostracizare). Ultimul răspuns ar putea fi extrem de ineficient întrucât s-ar putea să nu împiedice alte încălcări ulterioare ale proprietății sale. Primul răspuns ar putea fi privit ca o agresiune, mai degrabă decât ca o pedeapsă. Al doilea răspuns și, în special, al treilea sunt mult mai adecvate prin aceea că indică în mod clar o intenție de a pedepsi la nevoie, chiar cu asumarea unor anumite costuri de către acela care pedepsește.

Cred, însă, că ostracizarea sau evitarea constituie, în general, reacția cea mai importantă la încălcarea unei norme. Dacă în locul unor încălcări repetate ale proprietății vecinul ar fi încălcat o singură promisiune, atunci ruperea relațiilor ar putea părea reacția cea mai firească. Această teză este sprijinită și de ideea generală că normele sociale funcționează prin intermediul *emoțiilor* de rușine la nivelul aceluia care încalcă norma și dispreț la nivelul acelora care observă încălcarea (capitolul 8). Pentru că tendința spre acțiune a disprețului este evitarea, care produce de multe ori pierderi materiale pentru persoana ostracizată, există o legătură între reacția emoțională și impunerea unor sancțiuni. Însă, de multe ori, sancțiunile sunt mult mai importante ca vehicul pentru comunicarea emoțiilor decât sunt în sine. Mai mult, costul impunerii unor sancțiuni pentru *acela care sancționează* ar putea fi important, mai ales pentru comunicarea tăriei emoțiilor sale.

Teoria sancțiunilor cu privire la normele sociale întâmpină o problemă evidentă: ce îi motivează pe aceia care impun sancțiuni să pedepsească? Ei ce au de câștigat? În mod obișnuit, sancționarea este costisitoare și poate implica riscuri pentru acela care sancționează. Chiar dacă acesta nu renunță în totalitate la oportunitatea unor interacțiuni profitabile, exprimarea unui dezacord poate declanșa o reacție de furie și chiar violență din partea celui sancționat. Trebuie să facem aici distincția importantă între a dezaproba spontan și a face pe cineva de rușine în mod deliberat. În cel din urmă caz poate exista un recul care să îl determine pe cel sancționat să fie mai degrabă mânios decât rușinat. Chiar dacă dezaprobarea este spontană, cel supus dezaprobării poate interpreta gestul ca intenționat și poate reacționa în consecință. Din acest motiv, sancționarea este o chestiune riscantă. Prin urmare, ne putem întreba de ce, totuși, oamenii o practică.

Un răspuns ar fi acela că, în cazul în care nu pedepsesc, ei înșiși riscă să fie pedepsiți. Fără îndoială că așa ceva se întâmplă. Într-o societate în care există norme puternice care reglementează răzbunarea ne putem aștepta ca o persoană care nu evită pe cineva care trebuia să se răzbune, dar nu a făcut-o, să fie la rândul său ocolită de ceilalți membri ai societății. Printre școlari, un copil poate fi mai dispus să interacționeze cu un „tocilar" când colegii de clasă nu îl pot observa. Însă este puțin probabil ca un copil să suporte sancțiuni dacă nu se alătură unei găști care îi administrează o corecție unuia care s-a comportat prietenos cu un tocilar. Prin urmare, sunt slabe

șanse ca hărțuitorii terți să fie motivați de frica de pedeapsă. În experimente, problema ar putea fi abordată încercând să se stabilească dacă terții ar administra pedepse Respondenților care, acceptând oferte foarte mici în Jocul Ultimatumului, nu îi pedepsesc pe Propunătorii negeneroși. M-ar surprinde să aflu că i-ar pedepsi și aș fi chiar mai surprins să aflu că observatori de ordinul al patrulea i-ar pedepsi pe terții care nu pedepsesc. Mecanismul încetează să mai fie plauzibil la doar câteva noduri distanță de încălcarea inițială.

O explicație ceva mai parcimonioasă și adecvată a sancționării se bazează pe declanșarea spontană a disprețului și a tendinței spre acțiune asociată disprețului. Și mânia poate fi implicată din cauza distincției fluide dintre normele sociale și cele morale. De asemenea, este foarte probabil ca etalarea încălcării unei norme sociale să declanșeze mânie, nu doar dispreț, pentru că le comunică celorlalți oameni nepăsarea față de reacțiile lor. Cu toate că aceste tendințe spontane spre acțiune sunt ținute sub control de costurile și riscurile sancționării, uneori pot scăpa de sub control. Ostracizarea tocilarului care își ajută colegii de clasă cu temele este costisitoare, la fel cum a fost refuzul aristocraților din *Ancien Régime* să își lase fiicele să se căsătorească cu oameni de rând bogați. Eficiența economică poate fi afectată dacă un „gust" pentru discriminare ia forma refuzului de a face negoț sau de a angaja membri ai unui grup etnic disprețuit sau femei. De multe ori, astfel de comportamente reflectă funcționarea normelor sociale mai degrabă decât preferințele individuale idiosincratice, după cum o pun în evidență expresii ca „iubitor de evrei" sau „iubitor de negri", folosite pentru a-i condamna pe aceia care nu respectă norma.

Ce nu sunt normele sociale

Trebuie să distingem normele sociale de un număr de fenomene înrudite: norme morale, norme cvasimorale, norme legale și convenții. Cu toate că liniile de demarcație pot fi destul de fluide, există cazuri clare aparținând fiecărei categorii. Atât normele morale, cât și cele cvasimorale (capitolul 5) pot modela comportamentul, chiar dacă agentul crede că nu este observat de ceilalți. Dimpotrivă, rușinea care susține normele sociale este declanșată de disprețul perceput al celorlalți. Tendința spre acțiune corespunzătoare este aceea de a scăpa de privirile lor acuzatoare: de a te ascunde, de a fugi sau chiar de a-ți lua viața.

Normele legale sunt diferite de normele sociale prin aceea că sunt impuse de agenți specializați, care, de obicei, nu ostracizează, ci impun pedepse directe, chiar și în cazul experimentelor prin care s-a încercat determinarea unui simț al rușinii prin mijloace legale. Normele legale și cele sociale interacționează în variate moduri. De pildă, în 1990, unii legiuitori din Louisiana au cerut reducerea sancțiunilor penale aplicabile oamenilor care îi pedepsesc informal pe aceia care profanează simbolurile naționale. Chiar după ce un edict emis în 1701 a îngăduit nobilimii franceze să facă

negoț (chiar dacă numai angro), au trebuit să treacă mai bine de 50 de ani pentru ca aceștia să depășească normele sociale care interziceau o astfel de practică. În anumite comunități există norme sociale împotriva recursului la norme legale, iar în alte comunități oamenii se judecă imediat ce se comite un delict.

Convențiile sau echilibrele convenționale pot fi impuse, în principiu, prin simplul interes personal al agentului, fără să mai fie nevoie de acțiunea celorlalți. După cum s-a observat în capitolul 19, acestea sunt destul de arbitrare. În prima zi a unei conferințe, fiecare participant se așază mai mult sau mai puțin la întâmplare. În a doua zi a conferinței, convenția s-a creat deja: oamenii converg spre scaunele alese inițial pentru că acesta devine mecanismul de alocare evident (punctul focal). În a treia zi, convenția devine aproape un drept: mă supăr dacă un alt participant s-a așezat pe locul „meu". Însă, cu toate că normele sociale cimentează convențiile arbitrare și le sporesc șansele de a fi respectate, acestea nu sunt indispensabile. Printre newyorkezi există convenția de a sărbători ajunul Anului Nou în Times Square, însă, întrucât foarte puțini oameni știu dacă o anumită persoană vine sau nu, oportunitatea pentru a impune o sancțiune este foarte mică. Chiar dacă norma de a conduce pe partea dreaptă a drumului nu ar fi consolidată de normele sociale și legale, pericolele schimbării benzilor sunt atât de mari încât previn astfel de situații.

O categorie complexă este reprezentată de normele legale și politice nescrise, precum convențiile constituționale.[1] Acestea nu sunt, de obicei, impuse prin lege, cu toate că tribunalele pot ține cont de ele în actul de decizie. Sunt, însă, impuse de sancțiuni politice sau de frica de asemenea sancțiuni. Până în 1940, de exemplu, convenția constituțională din Statele Unite conform căreia un președinte nu poate executa mai mult de două mandate era impusă de credința că oricine ar încerca să încalce convenția ar pierde alegerile. Astfel de norme, care sunt numeroase, au ceva din normele sociale din moment ce sunt impuse de forța difuză a opiniei publice, și nu de vreo agenție specializată. Alte convenții politice se disting drept convenții de echilibru în jocuri repetate. În multe sisteme parlamentare, de exemplu, există convenția ca, atunci când o anumită administrație și-a terminat mandatul, documentele sale de uz intern să fie sigilate și puse la dispoziția istoricilor abia după câteva decenii. Cu toate că administrațiile ulterioare pot fi tentate să desigileze arhivele predecesorilor pentru a le folosi ca muniție politică, convingerea că un astfel de gest ar crea un precedent care i-ar face pe succesori să procedeze la fel este suficientă pentru a preveni o astfel de situație. Aceasta nu este o convenție în sensul din capitolul 19 din moment ce orice administrație ar prefera să devieze de la această practică dacă administrațiile ulterioare nu ar face-o.

[1] Expresia „convenție constituțională" se utilizează în relație cu două aspecte complet diferite ale constituțiilor: normele nescrise, care suplimentează constituția scrisă, și adunările constituante, care sunt de multe ori angajate în adoptarea unei constituții scrise.

Norme și externalități

Există norme împotriva acelora care impun mici externalități negative asupra celor mai mulți (capitolul 17). Oamenii care aruncă resturi pe joc în parc, care scuipă pe stradă sau urinează în lac încearcă de obicei să o facă fără să îi vadă nimeni. Chiar dacă nu le e teamă de sancțiuni, simplul gând că alții ar putea avea o părere proastă despre ei dacă ar face așa ceva îi poate descuraja de la asemenea acțiuni atunci când sunt observați. Normele de acest tip sunt utile din punct de vedere social în sensul tare că *toată lumea* are de câștigat de pe urma lor. Norma împotriva scuipatului în locuri publice este un exemplu foarte bun. Înainte să știm cum se răspândesc bolile infecțioase, scuipatul era o practică acceptată și reglementată prin facilități speciale, „scuipători". Odată înțeles mecanismul transmiterii bolilor, au început să apară semne cu „Scuipatul interzis" în multe localuri publice. Astăzi, norma este atât de înrădăcinată (cel puțin în anumite țări), încât astfel de semne nu mai sunt folosite.

În acest exemplu observăm cum apare norma și putem susține destul de încrezători faptul că a apărut *pentru că* era în interesul public. Pericolul era perceput, s-a creat o normă legală, iar norma socială i-a urmat. Mai discutabil este dacă percepția externalităților negative poate crea norme sociale fără pasul intermediar al unei intervenții publice. Simplul fapt că e nevoie de o normă și că există o percepție că norma lipsește nu este suficient pentru apariția normei. În țările în curs de dezvoltare nu există norme sociale care să limiteze dimensiunea familiilor. Normele sociale împotriva păscutului intensiv sau pescuitului excesiv nu au apărut spontan pentru a preveni tragedia bunurilor publice. Nicio normă nu reglementează utilizarea antibioticelor, cu toate că utilizarea lor în exces impune externalități asupra celorlalți prin dezvoltarea unor microorganisme mai rezistente. Normele împotriva ascultării muzicii pe o plajă publică sau împotriva utilizării telefoanelor mobile în sălile de concert (cred că) își au originea în acțiunile autorităților competente. Tot mai des descoperim că este nevoie de intervenție din afară pentru a-i opri pe oameni să le impună altora aceste externalități negative. În unele cazuri, precum în cazul normei care interzice scuipatul în spații publice, oamenii se pot abține, chiar dacă norma legală dispare sau încetează să mai fie pusă în practică. În alte cazuri, precum în China, unde există o reglementare care le impune familiilor să aibă doar un copil, pare puțin probabil ca, prin anularea reglementării, să se mențină respectivul comportament.

Grupurile mici pot fi în măsură să impună aceste norme fără intervenție externă. La locul de muncă există o normă destul de puternică împotriva acelora care depășesc planul de lucru pentru că se crede că eforturile lor ar putea scădea suma câștigată per unitate produsă. (În acest caz, externalitatea ia forma creșterii probabilității unei reduceri a sumei plătite.) Cu toate că patronatul poate dori să se angajeze în adoptarea unei politici referitoare la plata unor sume fixe per unitate produsă pentru a-și încuraja muncitorii să depună un efort mai mare, nu poate face o promisiune credibilă în

acest sens. Spărgătorii de grevă sunt, de cele mai multe ori, sancționați sever de colegii lor. Este, poate, destul de semnificativ că aceste două cazuri presupun o opoziție comună față de un adversar. Într-un „joc împotriva naturii", precum păscutul intensiv, solidaritatea nu apare atât de ușor pentru că parazitismul nu este considerat un act de *trădare*. Din acest punct de vedere, ar trebui să nu ne așteptăm la apariția spontană a unei norme împotriva evitării obligațiilor de muncă în firmele care oferă bonusuri pe echipă, din moment ce încălcarea acelei norme ar dăuna altor colegi fără să fie în vreun fel benefică pentru „inamic" (însă vezi capitolul 26).

Alte norme sociale ar putea să aibă în vedere externalitățile negative pe care un grup de oameni le impune altui grup. Norma împotriva fumatului, chiar și în locuri în care fumatul este încă permis prin lege, este un astfel de exemplu.[1] În multe societăți occidentale din ziua de azi oaspeții fumători se abțin adesea, chiar fără să întrebe gazda dacă au voie să fumeze. Ceva de felul unor „externalități de zgomot" fundamentează norma „Copiii ar trebui să fie văzuți, dar nu auziți." Sunt două moduri în care această dispoziție ar putea constitui o normă socială, și nu doar o formă de pedeapsă din partea părinților. În primul rând, copiii ar putea să-i ostracizeze pe alți copii care încalcă norma. În al doilea rând, părinții ar putea să-i ostracizeze pe părinții ai căror copii încalcă norma. În compartimentele de tren, aceia care vor să impună „externalitatea aerului curat" asupra celorlalți pierd, de obicei, competiția cu aceia care vor să impună o „externalitate a aerului îmbâcsit". Motivul ar putea fi acela că ferestrele închise sunt percepute drept o opțiune implicită și, prin urmare, un punct de referință normativ.

Norme și conformism

Unele norme sociale nu sunt nimic altceva decât dispoziții de *a nu-ți pune pielea în saramură*. Locuitorilor orașelor mici de pretutindeni li se va părea familiară „Legea lui Jante", scrisă (în 1933) de cineva care a scăpat:

Să nu crezi că ești *cineva*.
Să nu crezi că ești la fel de bun ca *noi*.
Să nu crezi că ești mai înțelept decât *noi*.
Să nu crezi că ești mai bun decât *noi*.
Să nu crezi că știi mai multe decât *noi*.
Să nu crezi că ești mai important decât *noi*.
Să nu crezi că *tu* ai vreo valoare.

[1] Cea mai importantă externalitate este provocată de inhalarea de fum (fumatul pasiv). Uneori se susține că fumătorii impun o externalitate negativă și asupra altor fumători care vor să se lase, dar nu pot rezista dorinței de a fuma provocată de semnalele vizuale declanșate la vederea celorlalți fumători.

Să nu râzi de *noi*.
Să nu crezi că este cineva interesat de *tine*.
Să nu crezi că *ne* poți învăța ceva.

Aceste norme pot avea consecințe sociale foarte grave. Ele pot descuraja persoanele talentate să își utilizeze talentul și pot duce la tratarea acestora drept vrăjitoare dacă își folosesc respectivele talente. Nici norocul nu e văzut cu ochi buni. În rândul populației Bemba din Rhodesia de Nord, se spune că a găsi un stup cu miere în pădure înseamnă noroc; să găsești două înseamnă foarte mult noroc; dar a găsi trei înseamnă vrăjitorie.

Codurile onoarei

Norme puternice și, de obicei, subtile pot reglementa comportamentul în dispute, vendete, dueluri sau în cazuri de răzbunare, în general. Normele definesc acțiunile care cer un răspuns sau o provocare, condițiile sub care și mijloacele prin care urmează să se desfășoare răzbunarea, precum și soarta aceluia care eșuează să respecte norma fundamentală. Începând cu ultima, eșecul de a te răzbuna determină adesea un soi de moarte civică, în care agentul este complet îndepărtat de relațiile sociale normale. În cadrul familiei, părerea sa este complet ignorată, iar dacă se aventurează în afara casei sale, este întâmpinat cu bătaie de joc sau chiar mai rău. Este o situație paradigmatică de dispreț, care determină o rușine intolerabilă.

Orice ar putea părea, indiferent cât de puțin, o insultă adusă onoarei agentului poate declanșa o revanșă. În Parisul prerevoluționar, Vicontele de Ségur, cunoscut ca fante în oraș, se distra scriind epigrame în versuri. Un rival gelos pe reputația vicontelui a scris o mică epigramă prin care își bătea joc în mod subtil de versurile lui Ségur. Ca răzbunare, Ségur a sedus-o pe amanta rivalului, iar apoi, când aceasta l-a anunțat că este însărcinată, i-a spus că doar a folosit-o pentru a se răzbuna pe rivalul său și că odată atins acel scop, nu mai are nevoie de ea. (Femeia a murit la naștere.) Vicontele s-a întors la Paris istorisind pățania oricui era dispus să îl asculte, neîntâmpinând niciodată dezaprobarea cuiva. Se pare că *Les liaisons dangereuses* este doar o palidă imitație a realității.

În Corsica secolului al XIX-lea existau patru circumstanțe care justificau sau impuneau răzbunarea: în cazul în care o femeie a fost dezonorată, în cazul ruperii unei logodne, în cazul în care o rudă apropiată a fost ucisă și când mărturia mincinoasă în tribunal a dus la condamnarea pe nedrept a unui membru al familiei. Un notar condamnat pentru crimă pe baza unei mărturii false a murit ulterior în închisoare. Fratele său a devenit bandit și, în decurs de câțiva ani, i-a ucis pe toți cei paisprezece martori ai acuzării. Toate acestea sunt exemple ale răzbunării de dragul *păstrării* onoarei. Sistemul onoarei include, însă, și acțiuni întreprinse cu scopul câștigării onoarei.

Montaigne se referă la „ce numesc italienii când vor să condamne acea bravadă impetuoasă a tinerilor ca *bisognosi d'honore*, «dornici de onoare»".

În sudul Statelor Unite oamenii reacționează foarte puternic la ceea ce percep a fi insulte, mult mai puternic decât cei din nord. Rata criminalității este mai mare în sud, iar oamenii aprobă mai ușor reacțiile violente la ofense. Într-un studiu ingenios, un complice al coordonatorului experimentului se izbește de un subiect în mod „accidental intenționat" și îl face pe acesta „tâmpit". Nivelul de cortizol de după (care reflectă reacția la incident) și nivelul testosteronului (care reflectă pregătirea pentru agresiune) cresc vizibil în cazul sudiștilor în comparație cu nordicii. Într-un alt experiment, subiecții continuă să se deplaseze pe un culoar de la locul incidentului spre un individ masiv (un complice), care se îndreaptă spre ei foarte decis. Culoarul fusese dinainte ocupat cu mese, astfel încât doar un singur om avea loc să treacă. Rezultatul este un Joc al Lașului. Cei din sud se apropie mult mai mult de persoana respectivă înainte să își piardă curajul (un metru și jumătate) în comparație cu nordicii (patru metri și jumătate).

Servesc, oare, codurile onoarei vreo funcție socială? Dacă da, funcția aceasta poate explica de ce există aceste coduri? Ideea că practicarea răzbunării este o formă utilă de control al populației este prea arbitrară pentru a fi luată în serios. O abordare alternativă, aceea că normele răzbunării furnizează un echivalent funcțional al reglementărilor legale impuse în cadrul unor societăți în care statul de drept este slab, pare și ea foarte neplauzibilă. Societățile mediteraneene și cele din Orientul Mijlociu, care au subscris la aceste norme, cunosc niveluri ale violenței și rate ale mortalității în rândul tinerilor care sunt mult peste cele întâlnite în altă parte.[1] După cum sugerează observația lui Montaigne citată mai devreme, normele răzbunării și codurile onoarei în care sunt incluse ar putea *aprinde mai multe incendii decât stinge*. De multe ori, vrajba între două familii poate duce mai degrabă la distrugerea unei ordini sociale, și nu la consolidarea ei.

Există și argumente conform cărora normele onoarei evoluează în societăți răzlețite, a căror principală activitate este păstoritul. În aceste societăți, folosirea violenței poate servi drept un instrument util, poate chiar indispensabil, pentru a descuraja furtul. Cultura onoarei în statele americane din sud ar putea fi explicată dintr-o astfel de perspectivă. Însă, dincolo de problemele generale reprezentate de explicația funcțională, analiza se lovește de dificultăți, pentru că și în alte locuri decât în mediile rurale

[1] S-ar putea obiecta că termenul relevant de comparație este nivelul violenței care ar rezulta în „starea de natură". Dacă definim acea stare printr-o preocupare excesivă față de interesul personal și prin absența oricărei agenții de tip stat, aceasta nu ar produce violență motivată de invidie, ciudă sau mânie. Perspectiva unei aproprieri prin forță de către cei puternici a bunurilor produse de cei slabi ar putea avea un efect de reducere a producției, împiedicând astfel violența să se producă.

codurile onoarei sunt foarte dezvoltate, de exemplu, la Curtea regilor Franței în secolele al XVII-lea și al XVIII-lea. Unii dintre aceia care se concentrează pe codurile onoarei prezente în aristocrația urbană, și nu pe acelea din rândul societăților pastorale, propun altă explicație funcțională: în absența unui război, nobilimea „are nevoie" de dueluri pentru a-și menține spiritul războinic. Dacă nu se furnizează și un mecanism prin care nevoia ar putea să își genereze propria satisfacție, atunci acest argument nu are nicio valoare. Toate aceste comentarii polemice nu presupun, însă, că eu am vreo explicație mai bună de oferit.

Normele de etichetă

Un alt set de norme sociale sunt acelea presupuse de regulile bunelor maniere sau de *etichetă*. Codurile de îmbrăcăminte, limbajul, comportamentul la masă și altele de acest tip sunt, de obicei, minuțioase și îi ostracizează pe aceia care omit chiar cea mai subtilă nuanță.[1] În toate societățile există o normă care reglementează distanța cuvenită pe care trebuie să o păstrezi față de ceilalți în societate. În cazul în care o persoană încalcă spațiul privat al alteia (în Statele Unite, undeva la 38 de centimentri) există riscul de a fi evitat ca necivilizat. Această normă este, însă, neobișnuită, în sensul că indivizii respectivi nu sunt de cele mai multe ori conștienți de existența și funcționarea acestei norme. Cele mai multe norme de etichetă sunt puternic codificate, adesea la propriu. Acestea nu sunt doar (în bună măsură) inutile, ci uneori și crude prin consecințele lor, ca în cazul în care o fetiță de cinci ani ajunge acasă în lacrimi pentru că prietenii ei și-au bătut joc de ea pe motiv că noul ei cărucior pentru păpuși *nu are frână*. În Parisul prerevoluționar, un tânăr ofițer bogat, dar fără rang, a încercat să intre neinvitat la un bal de la Versailles. „A fost tratat atât de sever, încât, în disperarea sa față de bătaia de joc cu care a fost tratat, într-un moment în care a fi ridiculizat era cel mai rău dintre rele, s-a sinucis când s-a întors la Paris."

Este paradoxal de ce aceste situații, în sine triviale, capătă o importanță atât de mare. Dezaprobarea disproporționată declanșată de o încălcare a etichetei se poate datora credinței nefondate că oamenii sunt „dintr-o bucată" (capitolul 10), prin urmare, dacă o persoană încalcă o normă neimportantă, sunt șanse mai mari să încalce și una importantă. De asemenea, încălcarea unor banale norme de etichetă poate fi percepută drept o manifestare a lipsei de considerație pentru ceea ce cred ceilalți oameni. Însă tot rămâne neexplicată existența în sine a acestor norme neimportante. Frapant nu este de ce există cutare sau cutare normă, ci de ce ar atașa oamenii o semnificație atât de puternică unor chestiuni în sine fără nicio valoare. Firește, odată ce normele

[1] În societățile aristocrate sunt permise uneori devieri foarte mari dacă sunt percepute ca devieri deliberate, și nu ca dovezi ale ignorării regulilor. Charlus al lui Proust este un astfel de exemplu.

sunt *percepute* drept importante, acestea pot *deveni* elemente importante în reglementarea comportamentului. În cazul în care cineva se prezintă la un interviu pentru un post într-o bancă îmbrăcat într-o jachetă de piele roz, acest lucru nu poate însemna decât că persoana încalcă deliberat norma socială sau că este în dezacord complet cu așteptările celorlalți. În ambele cazuri, comportamentul oferă un temei suficient de bun pentru a angaja pe altcineva.

Din nou, explicațiile funcționale sunt foarte răspândite. Se crede că regulile subtile de etichetă din rândul elitelor există pentru a îngreuna accesul acelora care nu sunt bineveniți în rândul lor prin imitarea comportamentului guvernat de reguli. Nu încape îndoială că aceste reguli au ca *efect* ținerea parveniților la distanță, însă acest lucru nu răspunde la întrebarea de ce există aceste reguli. După cum au descoperit mulți studenți susținători ai proletariatului, este foarte greu să intri în clasa muncitoare dacă nu te-ai născut în cadrul ei. De exemplu, în Norvegia anilor 1970, tinerii maoiști au descoperit că, dacă își bat joc de familia regală, se înstrăinează exact de clasa din care voiau să facă parte. Însă nimeni nu a sugerat că normele clasei muncitoare există *cu scopul* de a face mai greu pentru cineva din afară să se dea drept muncitor. Nici aplicat la normele elitei argumentul nu are mai mult sens.

Normele care reglementează utilizarea banilor

Pe lângă interdicțiile stipulate prin lege cu privire la folosirea banilor pentru a cumpăra bebeluși, voturi, note la școală, organe pentru transplant și (uneori) sex, există multe norme sociale care reglementează folosirea adecvată și, respectiv, inadecvată a banilor. Unele norme funcționează printre prieteni și vecini, precum norma funcțională în comunitățile suburbane care interzice să-l rogi pe alt vecin să-ți tundă peluza pe bani (capitolul 12). Dacă vecinii trebuie să coopereze pentru a ridica un gard, aceștia aduc contribuții de același tip, poate unul dintre ei pune la dispoziție munca, iar celălalt materia primă. Chiar atunci când ar fi mult mai eficient ca doar unul să facă munca și să fie remunerat de celălalt, norma împotriva tranzacțiilor financiare între vecini exclude o astfel de soluție. În Statele Unite, profesorii din universitățile private, cu excepția cazurilor în care sunt prieteni foarte apropiați, nici în vis nu ar îndrăzni să își întrebe colegii cu cât sunt plătiți.

Alte norme reglementează într-un mod întrucâtva surprinzător utilizarea banilor în relațiile dintre străini. Există, de pildă, o normă împotriva oferirii de bani celui care ocupă locul din față la coada pentru autobuz ca să cedeze locul respectiv.[1] Această normă este în mod evident ineficientă: dacă persoana căreia i se face propunerea este

[1] Dacă persoana respectivă spune că trebuie să ajungă rapid acasă pentru că are copilul bolnav, i s-ar putea permite să treacă în față, dar persoana căreia i-a luat locul urmează de obicei după aceasta. Din moment ce acest gest impune costuri asupra tuturor celorlalte persoane care așteaptă autobuzul, aprobarea lor trebuie obținută și ea într-un fel sau altul.

dispusă să mai stea o dată la coadă pentru bani, ambii agenți beneficiază de pe urma tranzacției și nimeni nu suferă. Conform lui Tocqueville, astfel de norme împotriva etalării averii în public sunt tipice pentru societățile democratice: „Îl vezi pe acest cetățean opulent? (...) Portul său e simplu, ținuta modestă. Luxul este adorat între cei patru pereți ai casei sale." Există chiar norme care reglementează discuțiile despre bani sau comentarea aspectului monetar al unei tranzacții. Meniurile duble, încă oferite cuplurilor în anumite restaurante, unul cu prețuri pentru bărbați și altul fără prețuri pentru femei, reflectă norma conform căreia galanteria ar trebui să nu fie întinată de preocupări pecuniare. Dacă duci o sticlă de vin la o petrecere, este de așteptat să îndepărtezi eticheta cu prețul.

Normele cu privire la băut

Dacă normele sociale sunt invariabil legate de sporirea bunăstării individului sau societății, ne putem aștepta ca ele să fie direcționate împotriva băutului în exces, care este perceput a avea consecințe dăunătoare pe termen scurt și lung. Într-adevăr, există un număr mare de norme de acest tip. Unele norme, de obicei legate de religie, impun abstinența totală. Islamul și anumite secte protestante impun interdicții totale asupra alcoolului. Dimpotrivă, normele seculare recomandă, de multe ori, băutul moderat. Norma italiană „Să nu bei niciodată între mese" are un efect dual: pe de o parte, limitează consumul total, dar, pe de altă parte, reduce rata de absorbție a alcoolului, amortizând efectul pe termen scurt al acestuia asupra corpului. În Islanda există norme care interzic consumul de alcool în fața copiilor și băutul în timpul pescuitului.

Normele împotriva alcoolului nu sporesc, însă, întotdeauna bunăstarea socială. Există norme care condamnă abstinența, cum există și norme care îndeamnă oamenii să bea mult. Printre indienii Mapuche din Chile, este criticat băutul de unul singur, ca și abstinența; se consideră că astfel de comportamente arată o lipsă de încredere față de ceilalți. Cultura franceză tradițională îi condamnă și pe cei complet abstinenți, și pe bețivi. În Italia, neîncrederea față de aceia care se abțin este exprimată printr-un proverb: „Să mă ferească Dumnezeu de cine nu bea." În cultura tinerilor din multe țări, abstinenții sunt supuși unei presiuni puternice și sunt ridiculizați. Există, însă, multe societăți în care băutul în exces este recomandat din punct de vedere social. În Mexic și Nigeria, calitățile de tip macho manifestate prin abilitatea de a ține la băutură sunt admirate foarte mult. În Rusia prerevoluționară băutul în exces era obligatoriu în subcultura reprezentată de tinerii ofițeri.

Când abstinența este condamnată și băutul în exces este obligatoriu din punct de vedere social, aceia care vor să se abțină trebuie să recurgă la subterfugii. În Suedia, o întrebare obișnuită este „Vrei niște lichior sau conduci?". Este atât de acceptată, încât alcoolicii care se abțin sunt adesea obligați să spună că sunt cu mașina pentru că, în acest mod, sunt scutiți de presiunea socială care, altminteri, i-ar obliga să bea un pahar

la insistența gazdei. Norma băutului poate fi contrabalansată doar de altă normă (aceea împotriva conducerii unei mașini în stare de ebrietate). Similar, s-a susținut că o alternativă pentru latino-americanii care vor să iasă din sistemul social care le impune ocazional ritualuri care presupun consumul excesiv de alcool este aceea de a se converti la protestantism. Din nou, norma băutului poate fi depășită doar de altă normă, care are, în acest caz, sprijinul religiei.

Acestea sunt cazuri în care normele sunt utilizate strategic. Însă oamenii pot să se comporte strategic pentru a ocoli normele. Unii chinezi din Antichitate considerau alcoolul ca fiind sacru și îl consumau în cadrul ceremoniilor sacrificiale; în cele din urmă ajunseseră să facă sacrificii de fiecare dată când voiau să bea. În Spania faptul de a nu bea pe stomacul gol este o interdicție culturală tacită, prin urmare băutura se servește cu mâncare. În ambele cazuri se poate observa reversul legăturii cauzale inițiale: în loc să se respecte norma băutului doar atunci când faci X, oamenii fac X de fiecare dată când vor să bea.

Normele privitoare la bacșiș

Bacșișul pentru anumite servicii este un fenomen deloc neglijabil. Există estimări conform cărora bacșișul în restaurantele din Statele Unite se situează între 5 și 27 de miliarde de dolari pe an; dacă adăugăm bacșișurile lăsate șoferilor de taxi, coafezelor și altora, suma ar fi mult mai mare. Estimările proporției din venit alimentată din bacșișuri sunt între 8% (presupunerea Fiscului) și 58% pentru chelnerii care servesc mese cu mai multe feluri de mâncare. În anumite contexte, bacșișul este greu de înțeles, în alte contexte nu. Dacă mergi la același coafor de fiecare dată când ai nevoie de o tunsoare, bacșișul te asigură că vei beneficia de servicii de calitate; același lucru este valabil pentru mesele luate în restaurantele preferate. În schimb, bacșișul în cazul întâlnirilor unice, cum ar fi un drum cu taxiul sau o masă într-un restaurant în care nu te aștepți să mai vii vreodată, sunt destul de greu de înțeles. Un astfel de comportament este greu de înțeles din mai multe puncte de vedere: nu poate fi menținut prin sancțiuni aplicate de cealaltă parte de-a lungul timpului și nici de sancțiuni ale terților la momentul întâlnirii. Dacă ești singurul pasager din taxi, alți oameni sunt rareori în măsură să știe dacă i-ai lăsat un bacșiș generos șoferului și nici oamenii din restaurant nu vor observa cât de mulți bani îi lași chelnerului.

S-a susținut că bacșișul este o manieră eficientă de a-i remunera pe chelneri. Este evident mai ușor pentru client să monitorizeze calitatea serviciilor decât pentru proprietarul restaurantului. Prin urmare, descentralizarea funcției de monitorizare și punerea recompensei în legătură cu performanța observată sunt moduri de a depăși „problema agentului principal" (cum să-i împiedici muncitori să nu muncească), fapt care afectează multe relații contractuale (capitolul 26). Prin urmare, bacșișul poate fi o parte a „contractului implicit" care are ca scop sporirea eficienței. Însă, după cum

a afirmat Sam Goldwyn, un contract nesemnat nu valorează nici cât hârtia pe care a fost scris. Argumentul prin care se încearcă explicarea normelor sociale, la fel ca multe alte argumente, este doar o chestiune care ține de un funcționalism fără probe. Ideea că proprietarii restaurantelor care interzic bacșișul sunt eliminați din competiția cu aceia care îl permit este în întregime conjecturală și, oricum, nu poate explica motivul pentru care clienții dau bacșiș în restaurantele din ultima categorie. De asemenea, în urma unei evaluări empirice, bacșișul nu pare să treacă testele de eficiență specifice. De exemplu, bacșișul nu pare să aibă o prevalență mai mare în acele ocupații în care monitorizarea se face mai ușor. Faptul că, de multe ori, chelnerii își pun bacșișul la comun subminează, de asemenea, argumentul eficienței.

Nu știu de ce există o normă care impune bacșișul pentru anumite profesii, dar pentru altele nu. Odată ce există o normă, însă, putem înțelege de ce oamenii plătesc acel bacșiș: pur și simplu, nu le place ideea că alții, cum ar fi un șofer de taxi dezamăgit, ar putea să îi dezaprobe, chiar dacă nu se așteaptă să se mai întâlnească vreodată cu acel individ. Nu este ceva necesar să fii obiectul privirii disprețuitoare a celuilalt. Poate fi suficient să știi sau să ai motive să crezi că celălalt simte disprețul. Pentru a da un alt exemplu, credința că ceilalți ar putea să te dezaprobe explică de ce mă abțin de la a mă scobi în nas pe peronul metroului când trece un tren fără să oprească, deși nu mai e nimeni pe peron.

De ce norme?

Importanța normelor sociale pentru reglementarea comportamentului și mecanismele cele mai apropiate prin care acestea funcționează sunt, după cum am spus, în general relativ bine înțelese. Nu cred, însă, că înțelegem foarte bine și originea lor. Acestea sunt două chestiuni complet diferite. În primul rând, care este originea, din punctul de vedere al evoluției, a emoțiilor corelate de rușine și dispreț, care susțin normele sociale? Cu alte cuvinte, de ce există norme sociale, de fapt? În al doilea rând, de ce există norme specifice în societăți specifice? Cum și de ce apar acestea? Cum și de ce dispar?

Un răspuns simplu la prima întrebare este că ne pasă foarte mult de ceea ce cred ceilalți despre noi. Căutăm aprobarea acestora și ne temem de dezaprobarea lor. Însă acest răspuns nu face decât să ridice aproape aceeași problemă: de ce ar trebui să ne pese de ceea ce cred alții despre noi? Firește, în anumite cazuri, consolidarea unei reputații poate fi un lucru util, care merită cultivat. Însă gândul că șoferul de taxi ar putea avea o părere proastă despre noi dacă nu îi lăsăm bacșiș este complet distinct de orice preocupări cu privire la reputație. De asemenea, din moment ce alții au o părere proastă despre noi deoarece că am încălcat o normă socială, explicarea normelor sociale prin dorința altora de a nu crede rău despre noi este, într-o anumită măsură, circulară.

Cu privire la ultima chestiune, cel mai obișnuit răspuns este că normele apar pentru a reglementa externalitățile. Ideea are o anumită valoare dacă adăugăm, după cum am argumentat că ar trebui să facem, faptul că normele sociale împotriva impunerii unor externalități negative asupra celorlalți sunt, de obicei, introduse de o autoritate externă. Există o normă socială generală de a respecta legea. Dacă am vedea amenzile ca prețuri și nu am percepe închisoarea mai stigmatizantă decât statul în spital, atunci nu ar exista o astfel de normă, dar, în general, aceste reacții la încălcarea legii nu se consideră că sunt echivalente cu alte poveri obiectiv egale.[1] Când legea interzice un comportament care impune externalități negative asupra celorlalți, norma socială a respectării legii poate deveni o normă împotriva respectivului comportament. Norma poate rămâne în vigoare chiar dacă legea care i-a dat naștere cade în desuetudine. Însă poate fi greu de făcut o distincție între un astfel de rezultat și apariția unui „echilibru bun" într-un Joc al Asigurării (capitolul 19). Dacă statul determină cooperarea pedepsindu-i pe trădători și, ulterior, înlătură aparatul care impune pedeapsa, oamenii pot continua să coopereze pentru că situația cea mai preferată a fiecărei persoane este aceea în care respectiva persoană și restul cooperează (nu există tentația unor acte parazitare).

În ce privește normele variate pe care le-am discutat, cum ar fi norma de a oferi bani pentru a cumpăra locul cuiva care stă la coadă la autobuz, normele de etichetă sau normele privitoare la bacșiș, este mult mai greu de găsit o explicație pentru apariția și rezistența lor în timp. Un mod de abordare propus de economiști este acela că rezistența normelor poate fi explicată ca un comportament de echilibru, iar apariția lor este o chestiune care ține de accidental și de istorie, lucruri despre care știința socială are puține de spus. Întrucât o premisă implicită a cărții de față este aceea că linia de demarcație dintre științele sociale și istorie este artificială și inutilă, nu pot fi de acord cu ultima teză. Iar cu privire la prima, am argumentat că normele sociale nu manifestă în mod obișnuit logica celui mai bun răspuns care caracterizează jocurile de strategie. Dacă observ – fără a fi eu însumi observat – că cineva încalcă o normă, cel mai bun răspuns, de obicei, nu este acela de a-l sancționa.

◈◈◈

Notă bibliografică

Capitolul de față este o extindere și (sper) îmbunătățire a interpretării pe care am dat-o normelor în *The Cement of Society* (Cambridge University Press, 1989) și, mai

[1] În Norvegia, conducerea mașinii sub influența alcoolului era sancționată cu trei săptămâni de închisoare. Unii oameni aduceau cu ei în celulă o lampă cu ultraviolete pentru a se bronza, încercând în acest mod să întărească povestea conform căreia au fost în vacanță.

pe scurt, în „Social norms and economic theory", *Journal of Economic Perspectives*, 3 (1989), pp. 99-117. Discuții semnificative cu privire la normele sociale pot fi găsite în J. Coleman, *Foundations of Social Theory* (Cambridge, MA, Harvard University Press, 1990), R. Ellickson, *Order Without Law and Social Norms* (Cambridge, MA, Harvard University Press, 2000). Am avut de învățat din aceste surse, însă niciuna nu m-a convins. Pentru o critică instructivă a lui Posner, vezi recenzia lui R. McAdams, *Yale Law Journal*, 110 (2001), pp. 625-690. Discuții utile despre normele constituționale nescrise sau convenții se pot găsi în două articole de J. Jaconelli, „The nature of constitutional convention", *Legal Studies*, 24 (1999), pp. 24-46 și „Do constitutional conventions bind?", *Cambridge Law Journal*, 64 (2005), pp. 149-176. Legea lui Jante este luată din A. Sandemose, *A Fugitive Crosses His Trail* (New York, Knopf, 1936). Rolul vrăjitoriei în sprijinul normelor împotriva acelora care își pun pielea în saramură sunt discutate în K. Thomas, *Religion and the Decline of Magic* (Harmondsworth, Penguin, 1973). Am discutat codurile onoarei și răzbunării în capitolul 3 din *Alchemies of the Mind* (Cambridge University Press, 1999). Povestea despre Vicontele de Ségur este luată din *Les mémoires de la Comtesse de Boigne* (Paris, Mercure de France, 1999), vol. 1, pp. 73-74. Aceste memorii sunt și sursa poveștii despre tânărul ofițer care s-a sinucis de rușine că a fost ridiculizat (*Ibid.*, p. 38). Normele de etichetă sunt subiectul lui P. Bourdieu în *Distinction* (Cambridge, MA, Harvard University Press, 1987), în care abordarea este puternic funcționalistă. Studiile experimentale despre „cultura onoarei" sunt descrise în R. Nisbett și D. Cohen, *The Culture of Honor* (Boulder, CO, Westview Press, 1996). Exemplele despre normele care reglementează consumul de alcool provin din cartea mea *Strong Feelings* (Cambridge, MA, MIT Press, 1999). Problemele studenților norvegieni care susțineau muncitorimea sunt minunat descrise într-un roman deosebit de amuzant, dar care, din păcate, nu este tradus în engleză, al lui D. Solstad, *Gymnaslærer Pedersens beretning om den store politiske vekkelsen som har hjemsøkt vårt land* (Oslo, Gyldendal, 1982). Pentru norme împotriva adresării de întrebări cu privire la câștigurile salariale, vezi M. Edwards, „The law and social norms of pay secrecy", *Berkeley Journal of Employment and Labor Law*, 26 (2005), pp. 41-63. Explicația bazată pe eficiență a normei bacșișului este oferită de N. Jacob și A. Page, „Production, information costs and economic organization: The buyer monitoring case", *American Economic Review*, 70 (1980), pp. 476-478. Este criticată în M. Conlin, M. Lynn și T. O'Donoghue, „The norms of restaurant tipping", *Journal of Economic Behavior & Organization*, 5 (2003), pp. 297-321, care propune o abordare mai apropiată de cea schițată aici.

Capitolul 23

FORMAREA DE CREDINȚE COLECTIVE

Tocqueville despre conformism

Mecanismele formării de credințe, pe care le-am examinat în capitolul 7, funcționează în bună măsură la nivelul individual, cu alte cuvinte, credințele nutrite de o persoană se datorează în mică măsură acelora nutrite sau exprimate de ceilalți. În capitolul de față voi discuta câteva mecanisme ale formării colective sau interactive de credințe. Pentru a ilustra această distincție, putem lua în considerare analizele lui Tocqueville despre conformismul american. Una dintre explicațiile pentru faptul că americanii au tendința de a avea aceleași idei este simplul fapt că au trăit în condiții asemănătoare. Din moment ce „oamenii egali în condiții (...) văd lucrurile din același punct de vedere, mințile lor au tendința naturală spre idei analoge și, cu toate că fiecare dintre ei poate fi diferit de contemporanii săi și își poate forma propriile credințe, toți ajung într-un mod inconștient și neintenționat să împărtășească un anumit număr de credințe comune". O altă explicație se bazează pe presiunea de a te conforma: „Cei mai mulți americani înalță o formidabilă barieră în jurul gândirii. În limitele stabilite, scriitorul este liber, dar vai de acela care are îndrăzneala să treacă dincolo de acele limite! Acela nu are a se teme doar de un autodafé, ci trebuie să înfrunte tot soiul de neplăceri și persecuții zilnice."

Acest din urmă pasaj sugerează că oamenii se conformează în exterior din cauza presiunii sociale, dar nu neapărat și în interior. După cum scria tot Tocqueville, dacă ai un punct de vedere care se abate de la regulă, „concetățenii tăi te vor ocoli ca și cum ai fi necurat. Și chiar și aceia care cred în nevinovăția ta te vor abandona de teamă să nu fie la rândul lor ostracizați." Alte pasaje sugerează că, de fapt, conformismul pătrunde până în suflet, astfel încât oamenii își dezvoltă în cele din urmă o credință sinceră în punctul de vedere al majorității. Sunt sugerate două mecanisme, unul „rece" sau cognitiv și unul „fierbinte" sau motivațional. Pe de o parte, „pare puțin probabil (...) că, fiind toți în egală măsură iluminați, adevărul nu va aparține celor mulți". Pe de altă parte, faptul că „legile politice americane sunt astfel concepute încât

majoritatea este suverană (...) îi sporește semnificativ influența inerentă asupra intelectului întrucât omul nu are un obicei mai înveterat decât acela de a recunoaște înțelepciunea superioară a asupritorului său".

Descoperiri experimentale

L-am citat pe Tocqueville *in extenso* (și îl voi mai cita pe parcursul acestui capitol) pentru intuițiile sale acute cu privire la aceste chestiuni. Problemele pe care le-a identificat – conformismul spre exterior *versus* conformismul spre interior și mecanismele cognitive *versus* cele motivaționale – sunt încă de actualitate. Pentru a le aborda, voi face mai întâi referire la unele experimente clasice referitoare la conformism.

În cel mai faimos experiment, subiecților li s-a cerut să indice care dintre trei linii A, B și C se află mai aproape de o anumită linie D. Există trei condiții: privat, public dublu și public simplu. În condiția privată, subiecții și-au enunțat răspunsurile când nimeni nu mai era de față, cu excepția coordonatorului experimentului. În acest caz, 99% au indicat că D se află cel mai aproape de B, lucru care sugerează fără dubii corectitudinea răspunsului. Însă, în cele două condiții publice, o minoritate semnificativă de subiecți dau răspunsuri diferite. În ambele condiții, subiecții răspund după ce o serie de persoane (complici ai conducătorului de experiment) indică unanim A ca aflându-se cel mai aproape. În condiția public dublu, în care subiecții răspund în prezența complicilor, cam o treime sunt de acord că A este mai apropiată.[1] În condiția public simplu, în care subiecții enunță răspunsul în particular după ce aud ce spun ceilalți, conformismul se diminuează, dar nu dispare.

Excesul de conformism în condiția public dublu este produs, cel mai probabil, de *frica de a nu fi dezaprobat*. Conformismul rezidual în condiția public simplu se poate datora *învățării* („nu se poate ca atât de mulți să se înșele") sau *reducerii disonanței*. Cea din urmă explicație pare cea mai plauzibilă. Este puțin probabil ca persoanele care s-au conformat majorității în particular să o fi făcut doar pe baza învățării raționale, dată fiind valoarea cognitivă redusă a opiniei majorității. Trebuie să fi existat și un factor motivațional în acest caz.

Un alt experiment întărește această interpretare. În acest caz, subiecții au o sarcină mai ambiguă, să detecteze distanța pe care o parcurge o sursă de lumină într-o cameră întunecată. Cu toate că sursa este fixă, subiecții izolați consideră că lumina parcurge

---

[1] În capitolul 5 am abordat modul în care observarea celorlalți poate declanșa comportamente asemănătoare cu ale acestora prin norma cvasi-morală a echității, în timp ce a fi observat *de către* ceilalți poate declanșa comportamente asemănătoare ca urmare a fricii de a fi dezaprobat. În cazul formării de credințe, a fi observat de ceilalți poate produce conformitate prin frica de dezaprobare, în timp ce conformitatea produsă de observarea celorlalți apare fie ca urmare a învățării, fie ca urmare a reducerii disonanței.

zece centimetri („efectul autokinetic"). Auzind un complice spunând că lumina s-a deplasat între 38 și 40 de centimetri, subiecții estimează că distanța parcursă este de aproximativ 20 de centimetri. Când doi complici estimează în jur de 40 de centimetri, subiecții consideră că distanța este aproximativ 35 de centimetri. Prezența unui complice duce la creșterea estimării cu aproximativ 10 centimetri, iar prezența a doi complici duce la o creștere cu încă 15 centimetri.

Într-un proces de învățare bayesiană (capitolul 11) mă pot baza pe ceilalți observatori pentru a-mi corecta percepția sau memoria. Estimările lor cu privire la un anumit fapt, cum ar fi distanța parcursă de lumină, pot fi folosite pentru a-mi modifica evaluarea inițială. *Cât de mult* mă vor afecta depinde de credințele mele cu privire la temeinicia percepțiilor acestora și de numărul observatorilor. În acest experiment, se poate presupune că subiecții atribuie aceeași temeinicie fiecărui complice. Indiferent cum se ajunge la această atribuire, modificarea estimării inițiale produsă de auzul estimării primului complice ar trebui să fie mai mare decât aceea produsă de aflarea estimării celui de-al doilea complice.[1] Însă acest lucru vine în directă contradicție cu descoperirile experimentale din moment ce enunțul celui de-al doilea complice a determinat o ajustare *mai mare* decât enunțul primului. Pare a exista un efect de reducere a disonanței produs de disconfortul de a te regăsi în dezacord cu majoritatea, care nu poate fi diminuat prin învățare rațională.

Cel de-al doilea experiment are și o altă trăsătură interesantă. Acesta s-a desfășurat pe mai multe „generații", iar complicii au fost înlocuiți treptat cu subiecți naivi. Astfel, la a doua generație a unui experiment cu doi complici, un complice este înlocuit de un subiect naiv dintr-un experiment de primă generație, iar, la a treia generație, celălalt complice este înlocuit de un subiect naiv dintr-o generație anterioară. În cele din urmă, toți participanții ajung să fie subiecți naivi expuși anterior fie complicilor, fie altor subiecți care au fost expuși unor complici și așa mai departe. Experimentul a fost conceput în așa fel încât subiecții nou incluși la nivelul fiecărei generații să vorbească după alți doi participanți. Inițiatorii experimentului au anticipat că estimările artificial de mari se vor menține indefinit, dar s-a dovedit că nu au avut dreptate. Cam după șase generații, în grupuri de trei persoane, și după opt generații, în grupuri de patru persoane, estimările converg spre 10 centimetri, adică distanța estimată în cazul în care avem de a face doar cu subiecți izolați. Credința în hainele noi ale împăratului nu a supraviețuit indefinit. Dacă anumite credințe culturale care sunt slab sprijinite de realitate se mențin în timp, acest lucru s-ar putea datora faptului că discrepanța dintre

[1] În exemplul numeric pe care l-am folosit în capitolul 11 pentru a ilustra învățarea bayesiană, fiecare nouă probă care confirmă o ipoteză aduce cu sine o creștere mai mică a probabilității în raport cu o probă anterioară. Această „diminuare a valorii marginale a informației noi" este un fenomen larg răspândit.

credință și realitate este greu de observat sau faptului că credințele sunt sprijinite de alte temeiuri. Folosirea loteriilor pentru a identifica locurile bune pentru pescuit sau vânătoare, după cum se procedează în anumite societăți, este posibil să fi supraviețuit datorită semnificației sale religioase.

Ignoranța pluralistă

La începutul capitolului am făcut o distincție între două motive pentru care, la un moment dat, oamenii ar putea nutri sau exprima credințe asemănătoare: pentru că sunt influențați de condiții asemănătoare (corelație) sau pentru că se influențează unul pe altul (cauzalitate). Un caz special al primului motiv este furnizat de numărul mare de exemple de descoperiri simultane, cum ar fi inventarea calculului diferențial și infinitezimal de către Leibniz și Newton în aproximativ aceeași perioadă de timp. Cu toate că nimeni nu poate identifica precis „condițiile asemănătoare", ideea calculului trebuie să fi „plutit în aer". Pentru un alt exemplu de apariție simultană a unor idei asemănătoare ne putem gândi la povestea cu noile haine ale împăratului, publicată de Hans Christian Andersen în 1835. În volumul al doilea din *Democrația în America*, publicat în 1840, Tocqueville aduce în discuție o idee asemănătoare pentru a explica stabilitatea aparentă a opiniei majorității:

> Timpul, evenimentele sau eforturile individuale depuse de spirite solitare pot, în anumite cazuri, să submineze sau să distrugă treptat o credință fără să existe vreun semn exterior că se întâmplă așa ceva. Nimeni nu combate credința sortită pieirii în public. Nu există nicio forță care se adună pentru a purta război cu o idee. Partizanii ei o abandonează tăcut unul câte unul, până când doar o minoritate o mai întreține. În această situație, dominația ei se menține. Din moment ce dușmanii ei continuă să fie în pace sau își fac cunoscute gândurile doar în secret, trebuie să treacă mult timp până când aceștia pot fi siguri că s-a produs o revoluție și, îndoindu-se de aceasta, nu pot face nicio mișcare. Aceștia veghează în tăcere. Majoritatea nu mai crede de mult, dar pare încă a crede, iar această fantomă a opiniei publice este de ajuns pentru a le îngheța sângele în vene potențialilor inovatori și pentru a-i reduce la o tăcere cuviincioasă.

Un pasaj din *Vechiul regim* al lui Tocqueville (1856) aduce un argument asemănător cu privire la religie. În timpul Revoluției Franceze, „aceia care și-au menținut vechea credință au început să se teamă că sunt singuri în loialitatea lor și, fiindu-le groază mai mult de izolare decât de erezie, s-au alăturat mulțimii fără să-i împărtășească credințele. Astfel, ceea ce era încă opinia unei părți a națiunii a ajuns să fie considerată opinia tuturor, iar apoi a început să pară o credință irezistibilă chiar și pentru aceia care i-au dat această falsă aparență."

În aceste pasaje, Tocqueville se referă la credințe pe care oamenii *declară* că le nutresc (sau că nu le dezavuează), și nu la credințele pe care le nutresc cu adevărat în mod

sincer. Din acest punct de vedere, analiza sa diferă de comportamentul manifestat în cazul experimentului cu lumina în mișcare și de condiția public simplu, în cazul experimentului cu compararea liniilor.[1] Distincția aceasta nu este, însă, indubitabilă. Am argumentat deja în mai multe locuri că nu este întotdeauna clar ce înseamnă „a crede" un anumit lucru. Chiar și în condiția public simplu, „credința" subiecților care au spus că A este linia cea mai apropiată se poate să nu fi fost foarte puternică. Este posibil, de exemplu, ca aceștia să nu fi fost în stare să parieze o anumită sumă pe afirmația respectivă. De asemenea, enunțarea unei credințe poate, în anumite circumstanțe, să determine tendința însușirii ei (capitolul 7).

Psihologia modernă a redescoperit intuițiile lui Tocqueville sub denumirea de „ignoranță pluralistă". În cazuri extreme, nimeni nu crede adevărul unei anumite propoziții, dar toată lumea crede că toți ceilalți îl cred. În cazurile mai realiste, cei mai mulți nu cred, dar cred că mare parte dintre ceilalți cred. Ambele situații sunt diferite de cazurile patologice, în care toată lumea declară public o anumită credință știind că nimeni nu o nutrește în particular. Comunismul a manifestat această *cultură a ipocriziei* într-un grad extrem, cel puțin în ultimul său stadiu, cel gerontocratic. Ignoranța pluralistă și culturile ipocriziei pot fi susținute de același mecanism, și anume frica de dezaprobare sau pedeapsa pentru exprimarea unor viziuni contrare majorității. Diferența stă în faptul că, în cazul ignoranței pluraliste, dezaprobarea este orizontală – aplicată de concetățeni care cred în mod fals că trebuie să îi ostracizeze pe aceia care se abat de la regulă dacă nu vor ca ei înșiși să fie ostracizați. După cum observă Tocqueville, cei care nu-i ocolesc pe rebeli pot fi ocoliți. Dimpotrivă, cultura ipocriziei funcționează prin aplicarea de pedepse de sus în jos sau pe verticală: aceia care nu își exprimă entuziasmul pentru îndeplinirea planului sau ura față de dușmanul de clasă își vor pierde foarte probabil locurile de muncă sau chiar mai rău. Pedeapsa pe verticală poate determina apoi măsuri orizontale, în cazul în care oamenii îi evită sau îi pedepsesc pe rebeli dacă vor să nu fie ei înșiși tratați ca rebeli.

Ignoranța pluralistă este diferită și de mecanismul subiacent în cazul sindromului martorului pasiv, observat în cazul uciderii lui Kitty Genovese. Într-o versiune (stilizată) a acestui caz, fiecare individ crede că pasivitatea celorlalți justifică propria sa pasivitate. Cauza nu poate să fi fost presiunea socială sau dorința de conformare la normele grupului, din moment ce 38 de martori erau prea izolați pentru a putea forma o comunitate. Mai degrabă, pasivitatea pare justificată de o inferență: întrucât nimeni altcineva nu pare să facă nimic, situația nu poate fi foarte serioasă. „Datele brute"

[1] De fapt, experimentul cu lumina în mișcare este public dublu, lăsând loc pentru nesinceritate. Însă natura ambiguă a sarcinii se poate presupune că facilitează adoptarea unor sincere sau cvasi-sincere credințe exagerate. Presupun că motivul pentru procedură este că studiul generațiilor succesive ar fi fost greu de implementat în condiția public simplu.

(țipetele femeii) sunt înăbușite de această inferență. Mă voi referi la acest mecanism în curând. Pentru moment, vreau doar să punctez faptul că situația nu presupune ignoranță pluralistă din moment ce nu există vreo discrepanță între credințele întreținute de fiecare în forul său interior și credințele pe care le atribuie celorlalți.

S-a arătat cum cultura consumului de alcool poate ilustra ignoranța pluralistă. În multe campusuri americane există o cultură a consumului de alcool, larg răspândită printre studenți, mai ales printre băieți. Cei mai mulți studenți nu sunt foarte obișnuiți cu consumul de alcool, dar îl practică deoarece cred, în mod greșit, că îl practică și mare parte din ceilalți.[1] Comportamentul lor se conformează cu ceea ce cred, în mod greșit, a fi atitudinea tipică în campus, mai degrabă decât cu atitudinile lor private. Un alt exemplu poate fi luat dintr-un experiment în care studenților li s-a spus să citească un articol scris deliberat într-un stil obscur, care face textul aproape de neînțeles, și li s-a cerut să spună cât de bine l-au înțeles și cât de bine cred că l-au înțeles ceilalți. Într-una dintre condiții, studenților li se oferă opțiunea de a cere ajutor coordonatorului experimentului; în altă condiție, li se spune că nu pot cere ajutorul. Chiar și în cazul primei condiții, niciun student nu s-a dus să ceară ajutorul pentru că procedura presupunea riscul că se vor face de rușine. Însă fiecare student pare să fi crezut că, deși nu cere ajutorul de frica de a se face de rușine, ceilalți nu îl cer pentru că au înțeles textul și nu au nevoie de vreun ajutor. Prin urmare, studenții în acea condiție aveau tendința de a crede că ceilalți au înțeles articolul mai bine decât l-au înțeles ei înșiși. Diferența s-a evaporat în cealaltă condiție. Putem emite conjectura că acest efect se datorează unui „sindrom al fratelui mai mare". După cum am semnalat în capitolul 18, toți suntem conștienți de propriile noastre frici și angoase interioare, dar, întrucât nu avem acces la viața interioară a celorlalți, avem tendința de a-i considera mult mai maturi și mai stăpâni pe sine.

În studiul despre consumul de alcool în campus s-a descoperit că, de-a lungul timpului, atitudinile private, credințele cu privire la atitudinea celorlalți și comportamentul ajung să se calibreze în rândul subiecților. Din aceasta decurge o întrebare cu privire la *stabilitatea* ignoranței pluraliste. Există două moduri prin care poate să dispară: când credințele false despre ceilalți devin adevărate sau când oamenii nu mai întrețin respectivele credințe. Dacă fiecare persoană adoptă credința pe care o atribuie (în mod fals) celorlalți, respectiva atribuire ajunge să se adeverească. Cel mai probabil, acest lucru se petrece prin reducerea disonanței, cauzată fie de disconfortul dezacordului cu majoritatea, fie de disconfortul cauzat de a spune un lucru și de a crede altul. Se pare că așa s-a întâmplat cu consumul de alcool în campus.

---

[1] Totuși, poate fi adevărat că atât cei mai mulți studenți nu beau, cât și că cei mai mulți prieteni ai celor mai mulți studenți consumă alcool, dacă studenții care beau au mai mulți prieteni decât aceia care nu beau.

Pe de altă parte, situația poate fi destrămată.[1] Să presupunem că 20% dintre membrii grupului manifestă în comportamentul lor faptul că nu împărtășesc credința respectivă, iar ceilalți 80% manifestă un atașament nesincer față de credința respectivă pentru că au nevoie de mai mult de 20% nonconformiști într-un grup pentru a deveni ei înșiși nonconformiști. Altfel spus, să presupunem că într-un grup de 100 există 20 de nonconformiști, 10 care ar fi dispuși să recunoască ce cred dacă cel puțin 25 au recunoscut deja, 15 care ar recunoaște dacă cel puțin 35 au recunoscut deja și 55 care s-ar alătura dacă 50 au recunoscut deja. După cum am mai spus, cultura majorității este stabilă. Ne putem, însă, imagina că 5 dintre cei mai conformiști indivizi pleacă sau mor și sunt înlocuiți de 5 nonconformiști. În acest caz, majoritatea s-ar dizolva. Cei 25 de nonconformiști ar crea condițiile pentru ca alți 10 să se alăture; ceilalți 35 ar mai atrage alți 15, formând astfel pragul pentru ca și ceilalți 50 să se alăture. În loc să ne referim la acest proces ca la *destrămarea conformismului*, putem folosi expresia *bulgărele de zăpadă al nonconformismului*. Vom observa o dinamică asemănătoare în cazul acțiunii colective (capitolul 24).

Conformismul poate fi destrămat în multe alte moduri. Copilul din povestea lui Andersen își face simțită prezența în experimentul comparării liniilor: când un singur complice enunță opinia adevărată că D se află cel mai aproape de B, conformismul dispare aproape în totalitate. Pentru alt exemplu, să ne gândim la credința răspândită atât în Anglia, cât și în Franța, înainte de Reformă, conform căreia regele putea vindeca scrofuloza prin atingerea bolnavului. Reformarea a subminat această credință întrucât atât catolicii în Franța, cât și anglicanii în Anglia s-au văzut obligați să explice de ce dovezile din țara vecină erau false. Însă recunoașterea posibilității unei erori colective pe scară largă s-a dovedit a fi periculoasă întrucât pretinsele dovezi false folosite pentru a susține credința în cealaltă țară nu erau foarte diferite de cele invocate în țara proprie.

Un alt mecanism al destrămării este publicarea unui sondaj de opinie. Înaintea referendumului din 1972 cu privire la intrarea Norvegiei în Piața Comună (cum se numea pe vremea aceea), guvernul, principalele partide din opoziție și cele mai mari ziare erau în favoarea intrării. Cu toate că referendumul pusese în evidență existența unei majorități populare împotriva intrării, fiecare oponent individual ar fi ajuns să creadă că face parte dintr-o minoritate restrânsă dacă sondajele de opinie nu ar fi arătat contrariul. Fără sondaje, rezultatul referendumului ar fi fost, cel mai probabil, altul. Unii dintre aceia care se opunea intrării nu s-ar fi dus la vot întrucât rezultatul ar

[1] Dispariția credințelor false în cadrul experimentului cu lumina în mișcare este tot o formă de clarificare, datorită faptului că în fiecare generație, subiecții folosesc propriile lor „date brute" pentru a ajusta distanța estimată într-un sens descendent prin comparație cu ce aud de la alții.

fi părut știut dinainte. De asemenea, mișcarea care s-a format pentru a-i convinge pe nehotărâți ar fi rămas de mici dimensiuni și fără influență. În perioada dintre introducerea votului universal și apariția sondajelor de opinie, ignoranța pluralistă în chestiuni legate de politică probabil că a fost destul de extinsă.

Zvonurile, fricile și speranța

O altă poveste a lui Hans Christian Andersen, *Nu încape îndoială*, pune în evidență modul în care „o singură pană se poate transforma foarte repede în cinci găini" prin exagerări succesive. Studiul formării zvonurilor și al propagării lor nu este, din câte știu, foarte avansat. Cele mai mari contribuții sunt aduse de unii istorici francezi (și anglo-americani) care au ca punct de plecare studiul deschizător de drumuri al lui Georges Lefebvre cu privire la „Marea Teamă" din 1789. Pe lângă Marea Teamă, aceștia au mai studiat și alte zvonuri, aproape toate false, cu privire la:

- reîntoarcerea lui Napoleon după cele două înfrângeri din 1814 și 1815
- restaurarea completă a vechiului regim (*Ancien Régime*)
- generalizarea socialismului în ajunul Revoluției din 1848
- invadarea masivă a Germaniei în martie 1848 cu muncitorii francezi sărăciți, care urmau să jefuiască, să incendieze și să ucidă
- o conspirație între doctori să „otrăvească populația"
- o conspirație între preoți și nobili să „înfometeze populația"
- iminența reducerii taxelor
- iminența creșterii taxelor
- luptătorii de gherilă care deschid focul asupra soldaților germani de pe acoperișurile caselor în timpul invaziei germane a Belgiei din 1914
- zeci de mii de soldați ruși se alătură trupelor aliate în august 1914

Făcând apel la aceste studii, cred că este posibil să identificăm anumite chestiuni generale și, poate, să tragem anumite concluzii. Voi începe prin a face o serie de observații empirice. În primul rând, ideea că o pană se poate transforma în cinci găini nu exagerează efectul de amplificare a zvonului. După insurecția muncitorilor din Paris în iunie 1848, doi bărbați observați pe marginea unui drum de țară au devenit zece, trei sute, șase sute, pentru ca, într-un final, să se audă că trei mii de „agitatori socialiști" (*partageux*) fură, dau foc și masacrează. Treizeci de mii de soldați au fost trimiși să contracareze amenințarea. În urma unei investigații, s-a descoperit că unul dintre cei doi era nebun, iar celălalt era tatăl său, care îl avea în grijă. În aceeași perioadă, un țăran a inventat o poveste pentru a speria un copil; imediat după aceea, mai bine de o mie de oameni se înarmaseră pentru a le veni de hac „tâlharilor" inexistenți.

În al doilea rând, este uneori posibil să identificăm cu o anumită precizie originea, viteza de propagare și mecanismul prin care zvonurile se propagă. Marea Teamă din 1789 a apărut în locuri diferite, aproape simultan, dar independent (coordonarea a fost asigurată de perioada recoltei), iar apoi s-a deplasat (cu o viteză estimată de patru kilometri pe oră), cuprinzând aproape toată țara. Zvonul unei invadări a Germaniei de către muncitorii șomeri în 1848 s-a deplasat cu o viteză comparabilă. În ambele cazuri, s-a estimat cât de mult a încetinit zvonul atunci când trebuia să se traverseze un munte sau în timpul nopții. În multe cazuri, acțiunile provocate de zvonuri erau și ele, după cum vom vedea imediat, inspirate de zvon. De multe ori, zvonul era propagat de indivizi care pretindeau, cu sinceritate sau nu, că se află în posesia unor informații: funcționari publici, vagabonzi, samsari, păstori, soldați care se întorceau de pe front. Zvonul se răspândea și prin tragerea clopotelor, care puteau fi auzite în satele vecine. Ulterior, ziarele și scrisorile au devenit surse importante ale zvonurilor. De multe ori, declarațiile publice de negare a zvonurilor mai mult au pus paie pe foc.

Pe scurt, Marea Teamă este explicată de Lefebvre prin aceea că „națiunea s-a speriat singură" (*le peuple se faisait peur à lui-meme*). Credința apropierii unor tâlhari a determinat mobilizarea de trupe, pe care unii țărani le-au confundat de la distanță cu tâlharii. Când dintr-un sat s-au auzit clopotele bisericii, detașamentele trimise de satele învecinate au fost confundate cu dușmanii. În 1848, un foc de avertizare tras de un tun într-un sat din Franța a fost interpretat de satele învecinate ca zgomot de război. Când zvonurile cu privire la o iminentă invadare a Germaniei de către francezii pauperizați au ajuns în Germania în martie 1848, drumarii de pe partea franceză a Rinului au trecut fluviul grăbiți pentru a se întoarce la casele și familiile lor. Alții, privindu-i de departe, e posibil să-i fi confundat cu invadatorii francezi.

În multe cazuri, substanța zvonului era existența unei conspirații împotriva poporului, orchestrată de guvern sau de elite. Un eveniment natural – o recoltă slabă, un șir de incendii și izbucnirea holerei – era atribuit unei acțiuni intenționate. (După cum am remarcat mai devreme, oamenii i-au perceput de multe ori pe cei din elite a fi motivați de răutate, și nu de propriul interes.) Același obicei mental a determinat de multe ori autoritățile să vadă o sursă comună intenționată în spatele a ceea ce, în realitate, avea de-a face cu zvonuri apărute independent. Din faptul că zvonuri asemănătoare apăreau simultan în mai multe locuri din țară autoritățile au inferat, corect, că acestea trebuie să fi avut o cauză comună. Dar, în loc să identifice drept cauză o situație obiectivă comună, cum ar fi o recoltă slabă, autoritățile au inferat, incorect, că se aflau în fața faptei unui agent intenționat. Zvonurile despre conspirații erau inseparabile de credința că zvonurile erau create de conspiraționiști.

Într-un comentariu despre tendința de amplificare a zvonurilor, Montaigne ne oferă ceea ce ar putea fi considerată prima analiză a micromecanismelor propagării zvonurilor:

> Distanța de la nimic la cel mai mic lucru este mai mare decât distanța de la cel mai mic lucru la cel mai mare. Când cei dintâi oameni se adapă din ciudățenia inițială și își răspândesc povestea, își pot da seama din opoziția pe care o trezesc că ceilalți găsesc dificil să le accepte povestea; aceștia sunt nevoiți să își cârpească povestea cu o bucată falsă de câlți. (...) La început, greșeala individuală produce una publică, apoi eroarea publică creează eroarea individuală. Trece apoi din mână în mână, întreaga țesătură este cârpită și remodelată, astfel încât martorul ei cel mai îndepărtat este mult mai bine informat decât martorul cel mai apropiat, iar ultimul care aude istorisirea este mai convins decât primul. Este un progres firesc. Pentru că oricine crede ceva este convins că are obligația să convingă și pe altcineva de faptul respectiv; iar pentru a face acest lucru nu-i este teamă să mai adauge, prin invenție proprie, orice poveste necesară pentru a depăși rezistența.

Abordarea lui Montaigne este întrucâtva binevoitoare, acesta neatribuindu-i aceluia care împrăștie zvonul nimic altceva decât dorința de a-i convinge pe ceilalți de ceea ce el crede a fi adevărat. În analiza asupra Marii Temeri, Lefebvre menționează și alte motive, mult mai obscure. Este posibil ca unii indivizi să fi fost motivați să exagereze pericolul pentru a nu fi acuzați de lașitate. Exprimarea neîncrederii putea aduce și acuzații că slujești contrarevoluția încercând să îi convingi pe oameni că nu au de ce să se teamă când pericolul bate la ușă, dar putea să implice, de asemenea, riscul ofensării amorului propriu al tuturor acelora care au dat alarma. În analizele zvonurilor din secolul al XIX-lea, s-a susținut că vagabonzii au tendința naturală de a răspândi zvonurile pe care ascultătorii lor ar vrea să le audă (o formă de autoînșelare indusă), cum ar fi întoarcerea lui Napoleon I pentru a le face pe plac susținătorilor săi sau boala lui Napoleon al III-lea pentru a le face pe plac opozanților săi. Similar, vânzătorii ambulanți ofereau cele mai senzaționale zvonuri care ar fi putut atrage o audiență cât mai largă. Acuratețea zvonurilor referitoare la luptătorii de gherilă belgieni (*franc-tireurs*) părea de necontestat odată ce fuseseră folosiți ca premisă pentru represalii sângeroase. Cum altfel, se puteau gândi germanii, ar putea fi justificată ferocitatea acestora? Iar când zvonurile au început să fie transmise de soldații răniți trimiși acasă de pe front, cine ar fi putut să le conteste?

Autorii care s-au aplecat asupra zvonurilor scot în evidență că acestea apar ca urmare a unei „conștiințe colective", care ia forma unor scheme preexistente care pot fi activate de cele mai nesemnificative evenimente. Credințele populare despre răutatea elitelor și tendința acestora de a conspira împotriva poporului se numără printre cele mai generale condiții care fac zvonurile credibile, pe lângă factorii specifici enumerați. Credințele despre nevoia insațiabilă a guvernelor pentru taxe mai mari și pentru mai mulți soldați au facilitat, de asemenea, acțiuni altminteri neutre, cum ar fi cercetările statistice care să declanșeze zvonuri conform cărora creșterea taxelor și a recrutărilor de soldați erau iminente. În 1914, credințele germanilor cu privire la natura Rezistenței

belgiene s-au format prin activarea schemelor despre luptătorii de gherilă create în timpul Războiului franco-prusac din 1870-1871. Aceste scheme au, de obicei, o anumită bază în realitate, chiar când credințele în sine pe care le inspiră nu au vreuna.

Am citat în repetate rânduri proverbul „Credem cu ușurință ceea ce sperăm și ceea ce ne inspiră teamă". Zvonurile pe care le-am discutat ilustrează ambele posibilități. Reîntoarcerea lui Napoleon, o modificare a taxelor odată cu o schimbare de regim și înrolarea soldaților ruși alături de forțele aliate, toate sunt martore ale puterii pe care o are speranța. (În cazul lui Napoleon, totuși, existau și oameni care se temeau de o eventuală întoarcere.) Însă un istoric al zvonurilor din Franța secolului al XIX-lea susține că, „în general, zvonurile sunt în mod sistematic mai mult pesimiste decât euforice", frica fiind mai des întâlnită decât speranța. Cu scopul de a verifica această teză, am putea încerca să numărăm episoadele de gândire deziderativă și antideziderativă pentru a vedea dacă, într-adevăr, cele din urmă sunt mai des întâlnite. Probabil că aceasta ar fi o întreprindere inutilă, atât pentru că nu am putea determina o mostră reprezentativă, cât și pentru că ar fi dificil să știm câtă importanță să atribuim fiecărui episod în parte.

O abordare mai promițătoare este aceea de a trasa o distincție între credințe și cvasi-credințe, cum am propus în Introducerea la Partea a II-a; diferența cea mai mare este aceea că doar primele sunt folosite ca premise pentru acțiune. Dacă ne raportăm la episoadele formării zvonurilor, se pare că, aproape fără excepție, *doar zvonurile bazate pe frică îi determină pe oameni să își modifice comportamentul.* În episoadele formării zvonurilor în Franța secolelor al XVIII-lea și al XIX-lea, țăranii recoltau grâul mai devreme pentru ca acesta să nu cadă pe mâna „tâlharilor", își vindeau recolta fără să păstreze sămânță pentru că le era teamă că va fi confiscată în întregime, își ascundeau obiectele de valoare când se zvonea că vor fi taxați suplimentar, se căsătoreau pentru a evita recrutarea și își făceau rezerve de sare când se zvonea că se va pune un nou bir pe sare. Zvonurile despre apropierea tâlharilor au avut o influență decisivă asupra Revoluției Franceze. Printre altele, au determinat atacul țăranilor asupra castelelor seniorilor, acțiuni care, la rândul lor, au generat decretul din 4 august 1789 care a abolit feudalismul. Răscoalele etnice sunt de multe ori declanșate de un atac iminent al celuilalt grup.[1] Gândirea antideziderativă generată de frică pare a avea o mare putere în influențarea comportamentului. Dimpotrivă, gândirea deziderativă generată de speranță pare a fi fost îmbrățișată pentru valoarea sa de consolare. Această asimetrie pare destul de solidă. Cu toate că nu dovedește că zvonurile bazate pe frică sunt mai numeroase, arată că acestea sunt mai importante pentru explicarea comportamentului.

[1] Rolul emoțiilor în răscoale este confirmat de tendința documentată conform căreia acestea apar mai frecvent când temperatura de afară e ridicată.

Cea mai mare excepție de la această asimetrie apare pe piețele financiare, unde zvonurile alimentate de speranță sunt de multe ori bazate pe baloane speculative. „Exuberanța irațională" a anilor 1990 a fost, cu siguranță, folosită ca bază pentru acțiuni în cele din urmă dăunătoare sieși. În același timp, piețele financiare manifestă și influența pesimismului irațional. Nu se știu, însă, prea multe despre mecanica și dinamica formării *interactive* a credințelor pe aceste piețe. Se pare că există o interacțiune complexă între aceia care acționează pe baza zvonurilor, aceia care acționează pe baza modificărilor de preț generate de zvonuri și aceia care acționează pe baza zvonurilor când acestea sunt susținute de modificări de prețuri. Cu toate că unii agenți acționează pe baza unor credințe care sunt raționale, date fiind probele pe care se bazează, o parte dintre aceste probe sunt rezultatul unor acțiuni iraționale întreprinse pe baza unor zvonuri nefondate.

Asimetria pare să fie restrânsă la formarea interactivă a credințelor. La nivel individual, gândirea dezideративă este, cu siguranță, la fel de capabilă să genereze comportamente ca și gândirea antidezideративă. Mai mult, putem observa fără prea multă greutate că, la nivel individual, ultimul mecanism este mai puțin întâlnit decât opusul său. Natura interactivă a zvonurilor pare să genereze, în moduri pe care le înțelegem foarte puțin, modele care se abat de la acelea pe care le observă când indivizii își formează credințele izolat unii de alții.

Cascade informaționale

Zvonurile pot apărea și prin formarea complet rațională de credințe, printr-un mecanism cunoscut sub numele de „cascadă informațională". Să presupunem că fiecare individ dintr-un grup are acces la o anumită informație privată referitoare la o anumită chestiune. Cu toții își formează credințele secvențial, fiecare bazându-se pe informațiile sale private *și* pe credințele exprimate de predecesori (în cazul în care există). De pildă, fiecare sătean ar putea avea probe private cu privire la prezența tâlharilor în vecinătate și ar putea folosi această informație, împreună cu ceea ce a auzit de la alții, pentru a-și forma o opinie pe care o dă mai departe. Votul prin strigarea catalogului poate avea o dinamică asemănătoare în cazul în care chestiunea în cauză implică doar credințe despre chestiuni factuale, și nu preferințe. Fiecare membru al unei adunări se va baza nu doar pe informația de care dispune, ci și pe ceea ce este dezvăluit de votul celor precedenți în catalog. Pentru un al treilea exemplu, putem să ne gândim la un referent științific pentru o revistă care află (dar nu știe și motivul) că alt referent a recomandat respingerea unei lucrări.

În aceste situații, oamenii utilizează concluziile procesului de formare de credințe al celorlalți ca materie primă indirectă pentru propria formare de credințe, fără să cunoască materia primă (informația privată) pe care au utilizat-o ceilalți pentru a-și forma concluziile. Prin urmare, se poate întâmpla ca indivizi raționali să ajungă să creadă

lucruri false, cu toate că aceiași indivizi ar fi ajuns la concluziile corecte dacă ar fi avut acces la „datele brute" ale predecesorilor, și nu doar la concluziile acestora. În exemplul de mai sus despre referentul științific, al doilea cititor ar fi putut, în urma lecturii raportului primului referent, să tragă concluzia că acesta este părtinitor și că se bazează pe raționamente false. Dacă, însă, cunoaște doar concluziile primului raport și faptul că prima revistă este foarte respectată, ar trebui să țină cont în mod rațional atât de opinia negativă a primului referent, cât și de propria sa evaluare. Dacă propria sa evaluare este favorabilă, dar numai ușor, poate recomanda în cele din urmă respingerea articolului. Un alt referent, cu o opinie personală puternic favorabilă, ar putea (tot rațional) să recomande respingerea dacă află că doi referenți anteriori au făcut acest lucru. Însă rezultatul ar putea fi suboptim (relativ la scopurile comunității de cercetători), întrucât ultimii doi referenți au favorizat publicarea, iar primul se poate să fi fost doar puțin împotriva publicării articolului.[1] Dacă recenzenții ar fi citit articolul în ordine inversă, concluzia ar fi fost diferită („dependența de cale").

◈◈◈

Notă bibliografică

Experimentele despre compararea liniilor, realizate prima oară de Solomon Asch, sunt descrise în orice manual de psihologie socială, de exemplu, cel al lui E. Aronson, *The Social Animal*, ediția a IX-a (New York, Freeman, 2003). Experimentul cu lumina în mișcare este descris în R.C. Jacobs și D.T. Campbell, „The perpetuation of an arbitrary tradition through several generations of laboratory microcultures", *Journal of Abnormal and Social Psychology*, 62 (1961), pp. 649-658. Pentru consumul de alcool în campus, vezi D.A. Prentice și D.T. Miller, „Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm", *Journal of Personality and Social Psychology*, 64 (1993), pp. 243-256. Pentru cercetarea referitoare la studenții care au citit articolul incomprehensibil, vezi D.T. Miller și C. McFarland, „Pluralistic ignorance: When similarity is interpreted as dissimilarity", *Journal of Personality and Social Psychology*, 53 (1987), pp. 298-305. Scenariul destrămării se bazează pe T. Kuran, *Private Truths, Public Lies* (Cambridge, MA, Harvard University Press, 1995). Observația despre efectul Reformei asupra credințelor în

[1] Diverse practici de reducere a conformării pot fi înțelese dintr-o astfel de perspectivă. Când evaluatorul intern la examenele unei universități trimite răspunsurile studentului unui evaluator extern (după cum se obișnuia la un moment dat în Norvegia), acesta nu trimite și nota pe care a dat-o. Din același motiv, când se caută o a doua opinie medicală, nu ar trebui să i se spună celui de-al doilea doctor diagnosticul dat de primul.

puterea vindecătoare a regelui i se datorează lui M. Bloch, *Les rois thaumaturges* (Paris, Armand Colin, 1961). Studiile formării de zvonuri pe care mă bazez le sunt datorate lui G. Lefebvre, *La grande peur de 1789* (Paris, Armand Colin, 1988), F. Ploux, *De bouche à oreille: Naissance et propagation des rumeurs dans la France du XIXe siècle* (Paris, Aubier, 2003), R. Cenevali, „The «false French alarm»: Revolutionary panic in Beden, 1848", *Central European History*, 18 (1985), pp. 119-142; M. Bloch, „Réflexions d'un historien sur les fausses nouvelles de guerre", *Revue de synthèse historique*, 33 (1921), pp. 13-35, și C. Prochasson și A. Rasmussen (editori), *Vrai et faux dans la Grande Guerre* (Paris, Editions la Découverte, 2004). O trecere în revistă detaliată și o analiză a zvonurilor în cazul răscoalelor etnice pot fi găsite în D. Horowitz, *The Deadly Ethnic Riot* (Berkeley, University of California Press, 2001). Pentru zvonurile de pe piețele financiare, vezi A.M. Rose, „Rumor in the stock market", *Public Opinion Quarterly*, 15 (1951), pp. 61-86. O introducere la mecanismul cascadelor informaționale o reprezintă S. Bikchandani, D. Hirshleifer și I. Welch, „Learning from the behavior of others: Conformity, fads, and informational cascades", *Journal of Economic Perspectives*, 12 (1998), pp. 151-170.

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

Capitolul 24

ACȚIUNEA COLECTIVĂ

Steagul viu

În *Lake Wobegon Days* (*Zilele orașului Lake Wobegon*), Garrison Keillor descrie Ziua Steagului în orașul său mitic. Herman, organizatorul paradei, a cumpărat o cantitate de șepci albastre, roșii și albe, pe care le-a distribuit fiecărui cetățean pentru ca locuitorii orașului să poată mărșălui pe străzi sub forma unui Steag Viu, în timp ce el urma să stea pe acoperișul Clădirii Centrale pentru a face o poză. Imediat după război, oamenii erau fericiți să se conformeze, dar mai apoi s-au răzgândit:

> Una dintre cauzele ranchiunei era faptul că nimeni nu putea vedea Steagul din care face parte; poza din ziar urma să fie alb-negru. Doar Herman și domnul Hanson aveau ocazia să vadă Steagul adevărat și câțiva băieți care erau prea scunzi pentru a putea servi la formarea Steagului. Oamenii doreau și pentru ei șansa de a urca pe acoperiș să vadă spectacolul.
>
> „Cum să urcați, când e nevoie ca voi să fiți jos?" spuse Herman. „Dacă vreți și voi să aruncați o privire, nu mai aveți la ce vă uita. Nu e suficient să știți că vă jucați rolul?"
>
> De Ziua Steagului, în 1949, tocmai când Herman spunea: „Asta e! Acum stați așa!", unul dintre cei roșii s-a desprins din mulțime, a urcat energic patru etaje până pe acoperiș și s-a aplecat să vadă spectacolul. Chiar și cu acea gaură rămasă în urmă, era o priveliște magnifică. Steagul Viu se întindea pe trei străzi. Un Steag perfect! Roșii erau minunați! Nu putea să-și ia privirea de la spectacol. „Coboară odată! Trebuie să facem poza!" urlă Herman la el. „Cum arată?" urlau cei de jos către el. „Incredibil! Nu pot să descriu ce văd" spuse acesta.
>
> Apoi toată lumea voia să vadă. „Nu!" spuse Herman, dar aceștia au supus chestiunea la vot și rezultatul s-a dovedit unanim. Unul câte unul, membrii Steagului Viu s-au suit pe acoperiș să admire priveliștea. Chiar *era* minunată. Lacrimile le podidiseră obrajii, te făcea să reflectezi asupra țării ăsteia minunate și asupra locului ocupat de orașul Lake Wobegon în toate astea. Ai fi vrut să zăbovești pe acoperiș toată după-amiaza și să absorbi totul. Așa că, pe măsură ce prima oră trecea și doar patruzeci din cei cinci sute urcaseră pe acoperiș, ceilalți

deveneau din ce în ce mai nerăbdători. „Grăbiți-vă! Nu mai căscați atâta gura! *Voi* ați văzut deja! Coborâți și lăsați și pe altcineva!" Herman începuse să trimită oamenii pe acoperiș în grupuri de câte patru, apoi de câte zece, dar, după două ore, Steagul Viu devenise Steagul Așezat, iar apoi începu să se erodeze pe măsură ce membrii săi, care deja se uitaseră, s-au gândit că ar fi bine să meargă acasă la cină, lucru care i-a înfuriat pe aceia care încă nu apucaseră să se uite.

„Încă zece minute!" urlă Herman, dar cele zece minute deveniră douăzeci, apoi treizeci, iar oamenii s-au furișat care încotro, ultimul privitor admirând un Steag care arăta de parcă ar fi fost ciuruit de un tun.

În 1959, fiii lui Knute și-au asumat răspunderea pentru Ziua Steagului. Herman le-a dat cutiile cu șepci. De atunci, frații Knute au reușit câteva Steaguri foarte bune, cu toate că prezența a fost destul de redusă în majoritatea zilelor. E nevoie de cel puțin patru sute de oameni pentru ca steagul să fie reușit. În unii ani, frații Knute au impus regula „nu priviți!", în alți ani au organizat o loterie. Într-un an au încercat să rezolve problema cu o oglindă uriașă ținută de doi bărbați deasupra acoperișului, doar că, odată ce oamenii și-au ridicat privirea să se vadă, Steagul dispăruse, firește.[1]

Wabegonienii se confruntă cu o problemă de acțiune colectivă, cu toate că una neobișnuită. Fiecare este tentat să urce pe acoperiș sau, în experimentele ulterioare, să se uite în oglindă. Dar, dacă toți fac asta, Steagul se *destramă* sau dispare și nu mai au la ce să se uite. Pentru a rezolva problema, aceștia recurg la tehnica clasică a coordonării: să impună o regulă de tipul „nu te uita!", să împiedice actele parazitare, să facă cu rândul, să organizeze o loterie. Prima regulă nu este foarte satisfăcătoare din moment ce nu are sens să faci un Steag pe care nu îl poate vedea nimeni.[2] Celelalte strategii asigură că numai unii oameni îl pot vedea, dar nu atât de mulți, ca să mai aibă ce vedea. Totuși, cer coordonare centralizată, susținută de organizații și de sancțiuni. Provocarea este aceea de a înțelege modul în care oamenii rezolvă uneori problemele impuse de acțiunea colectivă prin acțiune descentralizată.

Destrămarea și bulgărele de zăpadă

Dacă oamenii ar fi lăsați să facă ce vor, Steagul s-ar destrăma la propriu. Destrămarea în sens metaforic s-a dovedit a fi prezentă și în experimente care vizează bunuri publice. Subiecții sunt înzestrați cu o anumită sumă de bani și li se spune că, în cazul

---

[1] New York, Viking, 1985.

[2] Firește, Dumnezeu l-ar putea vedea. În anumite biserici medievale, numeroase capiteluri minunate se află atât de sus, încât nimeni nu poate vedea detaliile fără un binoclu care, la acea vreme, nu fusese inventat. Dar, din moment ce Dumnezeu le poate vedea, nu are nicio importanță. Se pare că wobegonienii nu au avut aceeași fervoare religioasă.

în care contribuie cu o parte pentru un fond comun, suma va fi înmulțită, iar suma astfel înmulțită se va distribui în mod egal tuturor subiecților, indiferent dacă aceștia au adus sau nu vreo contribuție. (Aceasta este o DP cu mai multe persoane, jucată de mai multe ori în rândul acelorași subiecți.) Dacă toți subiecții ar fi raționali și concentrați pe interesul personal, nimeni nu ar contribui cu nimic. De fapt, chiar și în interacțiuni anonime, subiecții încep prin a contribui în medie cu 40-60% din zestrea lor. În rundele ulterioare, contribuțiile scad progresiv, pentru a se stabiliza undeva la 10%. O explicație posibilă este aceea că majoritatea oamenilor sunt dispuși să contribuie dacă și numai dacă și ceilalți vor contribui la un anumit nivel. Nivelul poate varia de la individ la individ, unii au nevoie de o contribuție medie mare pentru a declanșa propriile lor donații, în timp ce, pentru alții, este nevoie doar ca unii să aducă

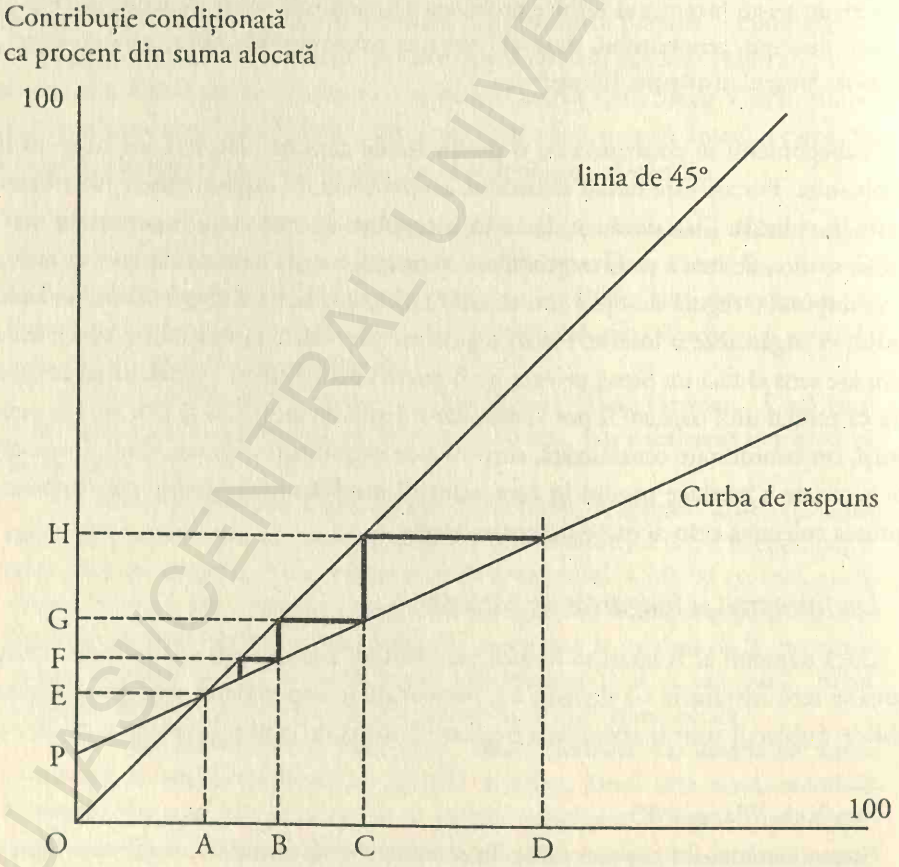

FIGURA 24.1

BCU IASI CENTRAL UNIVERSITY LIBRARY

o contribuție sau chiar nu este nevoie să contribuie cineva. Apoi destrămarea poate avea loc așa cum apare în Figura 24.1.

La început, oamenii cred că ceilalți vor contribui în medie OD. Unii indivizi sunt esențialmente egoiști și nu contribuie cu nimic, dar alții sunt mișcați de norma echității și aduc o contribuție. Suma totală a contribuțiilor lor este exprimată de OH. Această sumă mai mică (OH = OC) este folosită apoi ca bază pentru sumele așteptate în jocul următor. Unii dintre cei dispuși să contribuie când contribuțiile așteptate erau OD renunță acum, astfel încât contribuțiile rezultante scad până la OG. În rundele următoare, din ce în ce mai mulți contribuabili renunță în momentul în care nivelul contribuțiilor scade sub acela la care sunt dispuși să contribuie, până când procesul se stabilizează la OA = OE. În acest punct, contribuțiile actuale și cele așteptate coincid.

Am putea lua în considerare și fenomenul opus, „bulgărele de zăpadă", mai degrabă decât „destrămarea". Un exemplu este strângerea mulțimilor de oameni duminici la rând în Leipzig, înainte de dărâmarea Zidului Berlinului în 1989. Și în timpul revoluțiilor sau răscoalelor un număr inițial mic de participanți poate crește, transformându-se într-o mișcare suficient de puternică încât să dea jos guvernul. Înainte să încercăm să înțelegem astfel de fenomene, trebuie să explicităm cu atenție structura problemelor pe care le întâmpină acțiunea colectivă. Pentru moment, putem să ne bazăm pe caracterul secvențial al exemplelor date și să definim problema acțiunii colective în termenii relației dintre alegeri individuale și rezultate colective.[1]

Problema este puternic înrudită cu chestiunile pe care le-am discutat în capitolele anterioare. Problemele acțiunii colective apar în prezența externalităților pozitive sau negative care generează o DP care implică mai multe persoane sau o situație înrudită. Pentru a *defini* problema acțiunii colective mă voi limita la un subset de motivații care i-ar putea anima pe agenți în astfel de situații. Mai precis, se presupune că agenții sunt motivați doar de costurile participării și de beneficiile personale pe care le pot obține din rezultatul acțiunii colective. Un muncitor în grevă, de pildă, se presupune că se uită doar la riscul de a-și pierde locul de muncă și la salariul pe durata grevei, dar are perspectiva unui salariu mai mare (pentru el) dacă greva are succes. Dimpotrivă, s-ar putea să fim nevoiți să facem apel la un set mai larg de motivații pentru a *explica* alegerea strategiei cooperante în aceste situații.

---

[1] Voi continua prin a vorbi ca și cum alegerile individuale ar fi binare (indivizii fie cooperează, fie nu) și ca și cum rezultatele ar fi variabile în mod continuu (un bun public, cum ar fi aerul curat, ar putea fi furnizat într-o măsură mai mare sau mai mică). În realitate, indivizii nu diferă doar prin faptul de a contribui, ci și prin *cât* contribuie. Nu voi ține cont de acest aspect. De asemenea, unele bunuri publice sunt discrete sau „se prezintă în cantități mari". Dacă indivizii dintr-o comunitate fac lobby pentru a menține deschisă școala din localitate, ea fie va rămâne deschisă, fie se va închide. Această complicație poate fi rafinată, interpretând rezultatul drept *probabilitatea* variabilă în mod continuu a furnizării bunului public respectiv.

Furnizarea unei externalități pozitive implică costuri directe sau riscuri pentru agent. În experimentul referitor la bunurile publice, costul este acela al renunțării la o parte din proprietate. În cazul marșului din Leipzig, era acela al încasării unei bătăi din partea polițiștilor. În același timp, agentul furnizează un beneficiu material mic pentru aceia cu care interacționează. În cadrul experimentului, fiecare dintre ceilalți subiecți a primit o anumită pondere din contribuția sa. Și subiectul se bucură de un anumit beneficiu material (aceeași pondere cu ceilalți), dar, în mod evident, mai puțin decât costurile suportate. În Leipzig, fiecare membru al marșului devine un element în plus care atrage atenția polițiștilor (despre care putem presupune că sunt în număr constant) și diminuează probabilitatea marginală ca oricare dintre participanții la marș să fie bătut. Astfel, dacă ne limităm la costurile directe și la beneficiile directe ale cooperării, trădarea individuală este preferabilă cooperării individuale. În același timp, în ambele cazuri, cooperarea universală este preferabilă trădării universale. Dacă toți contribuie cu suma lor la fondul comun, fiecare va primi înapoi suma înmulțită. Dacă toți locuitorii din Leipzig ies în stradă, riscul pentru oricare dintre ei de a fi bătut este aproape de zero, în timp ce șansele de cădere a regimului detestat se apropie de 1.

Tehnologia acțiunii colective

Putem ilustra această situație prin intermediul unei diagrame. Într-un grup de $n + 1$ indivizi, Figura 24.2 indică modul în care recompensa pentru un anume individ variază ca o funcție a propriului său comportament și a comportamentului celorlalți. Comportamentul celorlalți este indicat pe axa orizontală, care măsoară numărul acelora care cooperează (printre ceilalți). Dacă individul este, de asemenea, cooperant,

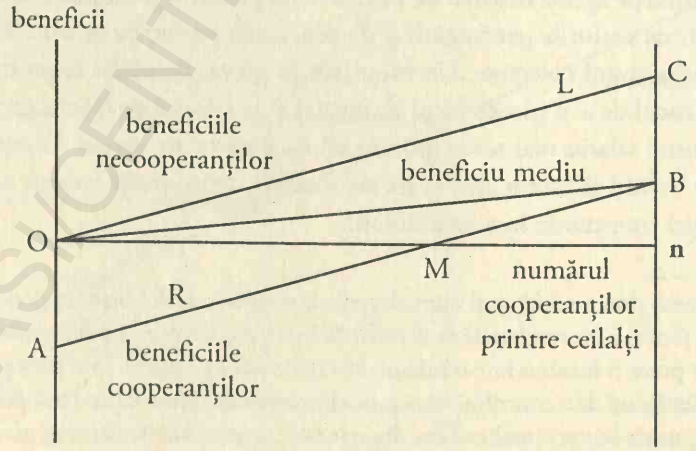

FIGURA 24.2.

Preluată din T. Schelling, *Micromotives and Macrobehavior* (New York: Norton, 1978)

utilitatea sa, măsurată pe axa verticală, este indicată pe linia R (AB) a diagramei. Dacă nu este cooperant, utilitatea sa este măsurată pe linia L (OC). Liniile L și R se intersectează cu axele verticale în ordinea care definește o DP obișnuită (de două persoane): rezultatul cel mai preferat este necooperarea unilaterală (parazitismul), cel mai bun rezultat următor este cooperarea universală, al treilea este necooperarea, iar rezultatul cel mai rău este cooperarea unilaterală (faptul de a fi exploatat). Ca și în situația cu două persoane, necooperarea este strategia dominantă din moment ce linia L este pretutindeni peste linia R. Spre deosebire de situația cu două persoane, putem defini un număr de cooperanți M care pot atinge un rezultat mai bun prin cooperare, chiar în prezența unor paraziți cărora le asigură un rezultat chiar mai bun decât al lor. Linia OB indică beneficiul *mediu* pentru toată lumea, atât pentru cooperanți, cât și pentru necooperanți, ca o funcție a numărului de cooperanți. Întrucât numărul agenților este constant, OB va reflecta și beneficiul *total* produs de cooperare.

Situația prezentată în Figura 24.2 reflectă un caz special. Ea presupune constanța costului cooperării, măsurat de distanța dintre curbele L și R. În alte cazuri, costul cooperării crește pe măsură ce mai mulți oameni cooperează. Pe măsură ce oamenii se alătură unor campanii prin intermediul radioului public, liniile telefonice se supraîncarcă și durează din ce în ce mai mult pentru fiecare dintre ei să intre în direct. Se poate chiar ca ultimii care se alătură[1] să reducă beneficiul mediu întrucât costul participării depășește suma beneficiilor pe care le generează pentru toți ceilalți (și pentru sine). De asemenea, se poate ca, inițial, costul să fie mai mare și să descrească ulterior. Pe măsură ce locuitorii din Leipzig se alăturau marșului, forțele guvernamentale erau obligate să se distribuie într-un mod mai risipit. Într-un astfel de caz, din motive asemănătoare, primii cooperanți vor fi înrăutățit situația (în sensul reducerii beneficiului mediu).

Figura 24.2 presupune, de asemenea, că beneficiile cooperării, date de linia L, sunt o funcție liniară a numărului de cooperanți. Fiecare cooperant nou adaugă aceeași sumă bunăstării tuturor. Beneficiile marginale în creștere pot fi ilustrate prin curățarea unei plaje de gunoaie: ultima sticlă înlăturată face, din punct de vedere estetic, o diferență mai mare decât penultima. Diminuarea beneficiilor marginale este, de asemenea, foarte frecventă. Un exemplu simplu este acela al unui apel la primărie cu privire la o groapă apărută în mijlocul unei străzi dintr-o comunitate urbană aparținând clasei medii: prima persoană care își face timp să sune ar putea ridica probabilitatea

[1] Atât aici, cât și în alte locuri, cuvinte ca „primul", „mijlociul" și „ultimul" se pot referi la momente la care se alătură cooperanți în succesiune, ca în constituirea unei mișcări revoluționare. Dar se pot referi și la acte simultane de cooperare, precum votul. A spune că ultimul votant adaugă puțin înseamnă că beneficiul creat în situația în care toți votează este aproximativ același cu beneficiul creat când aproape toată lumea votează.

reparării gropii la 0,4, cea de-a doua la 0,7, iar a treia la 0,8, a patra la 0,85, a cincea la 0,88 și așa mai departe. Uneori, atât primul, cât și ultimul contribuabil adaugă puțin, pe când cei din mijloc sunt mai eficienți. Un număr de revoluționari sau grevișți nu fac o diferență prea mare, iar când aproape toată lumea s-a alăturat mișcării, contează prea puțin dacă se alătură și aceia care încă nu și-au asumat respectivul angajament. În mișcările sociale, acest mod de desfășurare a lucrurilor este, probabil, cel mai întâlnit.

Beneficiile marginale ale cooperării pot fi chiar negative pentru o anumită mulțime de cooperanți. Dezarmarea unilaterală poate înrăutăți situația tuturor națiunilor dacă produce un gol de putere care poate îngădui invadarea respectivelor state, dezlănțuind astfel un război generalizat. Răscoalele izolate pot da autorităților un pretext pentru distrugerea răsculaților prezenți și a acelora potențiali. Invers, pot exista prea mulți cooperanți. Să presupunem că, în timp de război, toată lumea insistă să se alăture armatei, astfel încât industrii vitale pentru efortul de război rămân fără personal, iar războiul este pierdut în consecință. Dacă toată lumea vrea să ajute cu cina, prea mulți bucătari strică supa.

După cum reiese din aceste considerații, tehnologia acțiunii colective este diferită de la caz la caz. În cele ce urmează, mă voi concentra pe cazul indicat în Figura 24.3, care cred că este destul de des întâlnit în mișcările sociale care încearcă să obțină o schimbare a regimului politic. Primii participanți întâmpină costuri ridicate sau riscuri și produc beneficii restrânse celorlalți. De fapt, ar putea chiar să le facă rău altora în loc să le aducă beneficii. Contribuția lor netă este negativă. Ultimul participant produce, de asemenea, beneficii foarte restrânse. În unele cazuri, costul lor poate descrește, după cum am mai spus. În alte cazuri, toți luptătorii pentru o cauză anume pot înfrunta costuri considerabile și riscuri până la capitularea adversarului. Cei care s-au alăturat Rezistenței franceze în 1944 au adus puține daune naziștilor, dar toți au înfruntat riscuri considerabile pentru propria lor viață.

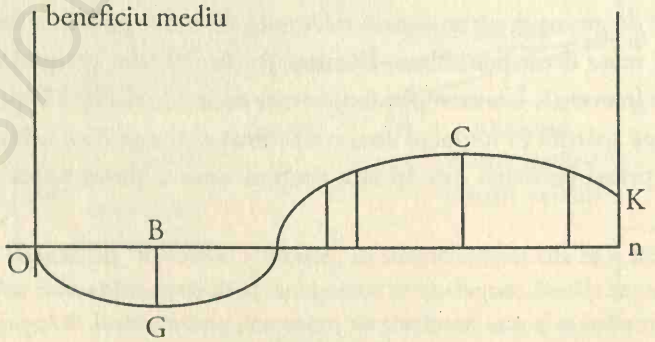

FIGURA 24.3

BCU IAȘI CENTRUL VECHI

Depășirea problemei parazitului

Am definit problema acțiunii colective în termenii motivațiilor egoiste orientate spre rezultate raționale. În mod obișnuit, acestea nu sunt suficiente pentru a genera cooperare din moment ce, indiferent de ceea ce fac ceilalți, individul este într-o situație mai bună dacă nu cooperează. Voi menționa în cele ce urmează două excepții de la această regulă.

Prima apare dacă *aceiași* agenți înfruntă *repetat* aceeași problemă de acțiune colectivă. Într-o interacțiune neîntreruptă de acest tip, cooperarea poate fi menținută de *strategia declanșării răspunsului cumplit*: cooperează cât timp toată lumea cooperează și trădează dacă un singur alt agent trădează o singură dată.[1] Există dovezi că, de pildă, cartelurile funcționează astfel. Decizia de a nu începe un război chimic în al Doilea Război Mondial se poate să se fi datorat fricii de represalii, dar, cel mai probabil, deficiențelor operaționale ale metodei. Aceste cazuri nu sunt, însă, obișnuite. De exemplu, muncitorii nu decid în fiecare an să își reînnoiască adeziunea la sindicat. Răscoalele și revoluțiile sunt mișcări unice, la care nu se aplică argumentul extras din interacțiunea repetată. Alte campanii, cruciade și lovituri au, de obicei, o structură prea rarefiată pentru a le îngădui cooperanților să-i identifice pe necooperanți. Pentru a explica modul în care apare cooperarea în astfel de cazuri trebuie să ne uităm în altă parte.

Cea de a doua excepție apare dacă există un agent exterior, în mod obișnuit, o organizație care poate impune pedepse selective asupra acelora care nu cooperează și/sau recompense selective pentru cooperanți. Statul îi poate pedepsi pe indivizii care impun externalități negative asupra celorlalți, de exemplu, asupra acelora care fumează în public sau scuipă pe stradă. Sindicatele pot oferi vacanțe mai ieftine membrilor. Mișcările revoluționare pot promite activiștilor poziții centrale în guvernele postrevoluționare. Aceste „soluții" la problema acțiunii colective se bazează pe o impunere centralizată, și nu pe acțiunea descentralizată a indivizilor pe care chestiunea îi privește direct. Trebuie, însă, să observăm că existența unei organizații capabile să modifice stimulentele agenților sociali poate ea însăși să fie rezultatul unei acțiuni colective de succes. De asemenea, promisiunile de a-i recompensa pe activiști după victoria unei mișcări pot fi credibile sau nu. În multe cazuri, de la Henric al IV-lea, în 1594, până la Charles de Gaulle, în 1945, liderii carismatici și-au dezamăgit adepții numind sau

---

[1] Spre deosebire de DP iterată, acest enunț se referă la o situație de durată indefinită. Dacă, într-un anumit caz, strategiile declanșatoare formează un echilibru, acest lucru depinde de structura recompensei cardinale și de rata deprecierii timpului din partea agenților. În principiu, și alte strategii pot genera un echilibru cooperant, dar numai strategiile declanșatoare au calitatea de a furniza un punct focal.

păstrând în poziții-cheie oficiali ai regimului înfrânt. Militanții nu sunt întotdeauna buni administratori.

În situațiile care implică acțiune colectivă unică și descentralizată, cooperarea nu apare dacă agenții sunt pe deplin raționali, dacă toți au motivații strict egoiste și orientate spre rezultate și dacă aceste două fapte sunt binecunoscute. (Această afirmație este un adevăr logic, nu un enunț empiric.) Referindu-ne la Figura 24.3, putem identifica un număr de motivații care pot declanșa și menține comportamentul cooperant. În scenariul stilizat pe care îl voi schița, aceste motivații se manifestă succesiv, fiecare dintre ele (cu excepția celei dintâi) bazându-se pe prezența agenților motivați de precedentele. Deși acest scenariu particular nu se aplică în toate cazurile, cred că una dintre trăsăturile sale are validitate universală. Acțiunea colectivă de succes, atunci când o putem observa, nu este opera unor indivizi cu motivații identice, ci, mai degrabă, este rezultatul unor *motivații mixte* – atât la nivel individual, cât și între indivizi.[1] În cele ce urmează, voi identifica o serie de astfel de motivații despre care cred că sunt prezente în proporții variate în multe situații care implică acțiune colectivă.

Cooperarea are nevoie de indivizi care nu sunt motivați strict de costurile și beneficiile personale. Acești indivizi pot fi împărțiți în două categorii, pe care le voi numi *utilitariști desăvârșiți* și *utilitariști dezinteresați*. Prima categorie va coopera dacă și numai dacă, prin contribuția lor, crește beneficiul mediu. În Figura 24.3, aceștia necesită ca numărul cooperanților să fi ajuns deja la un punct OB, la care acțiunea lor are consecințe utile. A doua categorie de indivizi va coopera dacă și numai dacă, prin contribuția lor, va crește beneficiul mediu, fără să se țină socoteală de costurile pentru ei.[2] În unele dintre exemplele pe care le-am dat mai devreme, cooperarea are efecte negative doar întrucât costurile pentru cooperator depășesc suma totală a beneficiilor. Utilitaristul dezinteresat nu va fi descurajat dacă acțiunea sa are efecte negative în acest sens. Se va abține, însă, de la a coopera dacă acțiunile sale aduc daune directe celorlalți. Mulți dintre aceia care vor să pornească mișcări sociale s-au găsit într-o situație de acest tip. După cum am remarcat deja, acestora le poate fi frică să nu provoace represalii din partea angajatorilor sau a autorităților, măsuri care le-ar putea fi dăunătoare membrilor grupului, și nu doar lor personal.

Dacă nu trăiesc o astfel de problemă, utilitariștii dezinteresați pot fi destul de numeroși pentru a ridica un număr de cooperatori la OB, unde intră în joc și utilitariștii desăvârșiți. Dacă utilitariștii dezinteresați sunt puțini sau dacă situația pe

---

[1] Pentru a nu complica prea mult discuția, voi lua în considerare doar variațiile în rândul indivizilor.

[2] Asimetria este surprinsă de câteva proverbe: „Prietenia înseamnă să uiți ce dăruiești și să ții minte ce primești" și „Nu uita niciodată un serviciu care ți se face, uită-l imediat pe cel făcut de tine".

care am descris-o îi descurajează de la a acționa, este nevoie de *cooperatori necondiționați* pentru a crea masa critică OB, care îi va pune în mișcare pe utilitariștii desăvârșiți. Acești actori vin în diverse forme – pot fi kantieni, sfinți, eroi, fanatici sau pot fi doar un pic nebuni. Au, însă, în comun faptul că nu acționează nici în funcție de consecințele așteptate ale acțiunilor lor, nici în funcție de numărul celorlalți cooperatori. De exemplu, naziștii fanatici s-au alăturat NSDAP într-un moment în care perspectivele erau cel mult o nălucă. Dimpotrivă, utilitariștii acționează în funcție de consecințele așteptate ale acțiunilor lor. Cu toate că așteptările lor pot fi în sine o funcție a numărului celorlalți cooperatori, acțiunea lor nu este declanșată direct de acel număr (capitolul 5). În cazul în care curba OGCK din Figura 24.3 nu scade mai jos de 0, ambele tipuri de utilitariști ar acționa ei înșiși ca forță motrice.

O altă categorie de actori, însă, nu ar acționa niciodată ca și cum ar fi o forță motrice. Aceștia sunt indivizii ale căror motivații sunt declanșate de observarea cooperării celorlalți sau de cunoașterea faptului că sunt observați de alți cooperatori (capitolul 5). Un prim subset al acestui grup conține indivizi motivați de norma cvasimorală a echității: nu este drept pentru noi să rămânem pe margine în timp ce alții își asumă riscuri pentru o cauză comună. Această motivație poate varia în intensitate, în sensul că un număr diferit de cooperatori poate declanșa cooperarea diverșilor indivizi. Aceia a căror cooperare este declanșată de un prag scăzut pot, prin urmare, să contribuie la declanșare pentru cei cu un prag mai ridicat. Un al doilea subset conține indivizi motivați de norme sociale. În cazul în care necooperanții pot fi identificați și supuși unei ostracizări sociale, după cum se întâmplă de obicei la locul de muncă, cooperatorii îi pot face să se alăture prin intermediul rușinii. Și rușinea are grade. Unii pot fi convinși să se alăture printr-o simplă remarcă sau printr-un act de evitare, în timp ce alții se alătură doar dacă sunt supuși unei presiuni din partea unui număr semnificativ de membri ai grupului, inclusiv (sau, poate, în special) din partea acelora care s-au alăturat, la rândul lor, în urma rușinii.

O ultimă categorie este reprezentată de aceia care se alătură unei mișcări pentru „beneficiile procesuale" – pentru că, personal, li se pare distractivă sau interesantă. Pentru unii oameni, faptul de a mărșălui și a cânta pe stradă sau gustarea oricărei forme de distracție oferite poate fi un lucru dezirabil în sine, indiferent de cauză. Pentru alții, alăturarea la o mișcare poate fi o ocazie de a-și „înălța conștiința" sau de a-și „clădi caracterul". Cu toate că stările dorite sunt produse esențialmente secundare (capitolul 4), ceea ce înseamnă că este foarte probabil ca aceia care le caută direct să fie dezamăgiți, aceștia se pot, totuși, alătura din acest motiv. Întrucât acești indivizi nu pot fi deosebiți la exterior de aceia din celelalte categorii, și ei pot servi ca element declanșator al adeziunii acelora cu un prag de adeziune ridicat.

Această tipologie a motivelor nu oferă nicio explicație pentru acțiunea colectivă de succes, ci doar un cadru pentru explicație. Există multe amestecuri de motivații

care pot crea condițiile pentru efectul bulgărelui de zăpadă. În unele cazuri, mișcarea poate să nu fie capabilă să recruteze suficient de mulți cooperatori dacă primii susținători sunt într-un număr prea mic pentru a-i atrage pe utilitariști sau dacă forțele motrice și utilitariștii sunt prea puțini, luați împreună, pentru a-i atrage pe aceia a căror adeziune este condiționată de adeziunea altora. În alte cazuri, procesul poate să nu se declanșeze deloc pentru că nu există forțe motrice. Rezultatul depinde de distribuția motivațiilor la nivelul populației și de tehnologia acțiunii colective. Depinde, de asemenea, și de organizare și de lideri, în moduri pe care le-am ignorat aici pentru a mă putea concentra asupra cazurilor pure ale alegerilor descentralizate.

◈◈◈

Notă bibliografică

Experimentul legat de binele public și destrămare este descris de E. Fehr și U. Fischbacher, „The nature of human altruism", *Nature*, 425 (2003), pp. 785-791. Pentru efectul bulgărelui de zăpadă în Europa de Est, vezi R. Petersen, *Resistance and Rebellion: Lessons from Eastern Europe* (Cambridge University Press, 2001). Această lucrare descrie și modul în care unele motivații pot coexista la nivelul aceluiași individ, nu doar (cum e cazul în textul meu) la nivelul unor indivizi diferiți. Am discutat unele varietăți ale tehnologiei în cartea mea *The Cement of Society* (Cambridge University Press, 1989); altele sunt discutate în G. Marwell și P. Oliver, *The Critical Mass in Collective Action* (Cambridge University Press, 1993). O trecere în revistă elegantă este făcută de T. Sandler, *Collective Action* (Ann Arbor, University of Michigan Press, 1992). O aplicație care ține cont de faptul că agenții individuali pot fi diferiți atât în resursele lor, cât și în motivațiile lor apare în J. Bowman, *Capitalist Collective Action* (Cambridge University Press, 1993). Cooperarea prin mecanismul declanșării răspunsului celui mai cumplit la nivelul cartelurilor este analizată în C. Lipson, „Bankers' dilemmas: Private cooperation in rescheduling sovereign debts", în K.A. Oye (ed.), *Cooperation Under Anarchy* (Princeton, NJ, Princeton University Press, 1986). Cazul negustorilor newyorkezi de diamante, discutat în capitolul 23, este relevant și aici. O abordare a acțiunii colective care scoate în evidență relațiile verticale între lideri și adepți este propusă de S. Kalyvas, *The Logic of Violence in Civil War* (Cambridge University Press, 2006).

Capitolul 25
DECIZIA COLECTIVĂ

De multe ori, membrii unui grup – de la familie până la societate ca întreg – trebuie să reglementeze chestiunile de interes general luând decizii care sunt obligatorii pentru toți. Putem să aducem din nou în discuție problema reglementării consumului de apă în perioade de secetă (capitolul 5). Uneori, această problemă de acțiune colectivă poate fi rezolvată prin decizii descentralizate, printr-o combinație de norme morale, cvasimorale și sociale. De multe ori, însă, consiliul local este nevoit să limiteze alimentarea cu apă sau să reducă consumul prin interzicerea anumitor utilizări, precum întreținerea peluzei sau umplerea piscinelor. Când acțiunea colectivă eșuează, poate fi nevoie de o decizie colectivă.

Putem da un alt exemplu, și anume practica votului în alegerile naționale. Alegerea între a vota și a sta acasă este o problemă clasică a acțiunii colective. Știind că faptul de a-ți face vocea auzită prin vot nu influențează deloc rezultatul, interesul fiecărui cetățean este acela de a absenta. Însă, dacă toată lumea ar rămâne acasă sau dacă votul s-ar situa la un nivel foarte scăzut, democrația însăși ar fi în pericol de a fi înlocuită de dictatură sau oligarhie împotriva interesului tuturor cetățenilor. În multe democrații, votul atinge niveluri cu adevărat respectabile, între 50 și 80%, ca rezultat al acțiunii descentralizate a cetățenilor. Unii cetățeni s-ar putea întreba: „Ce s-ar întâmpla dacă toată lumea ar lipsi de la urne?" Alții și-ar putea spune: „Din moment ce majoritatea oamenilor se deranjează să voteze, este firesc să fac și eu același lucru." Alții ar putea să gândească: „Deși impactul votului meu asupra viabilității democrației este mic, el devine important dacă este înmulțit cu numărul mare de cetățeni pe care îi afectează." Într-un sat de mici dimensiuni ar putea exista temerea că, „dacă stau acasă, vecinii mei vor observa și își vor exprima dezaprobarea".

Dacă aceste motivații, singure sau luate împreună, se dovedesc prea slabe, votul ar putea scădea la un nivel dezastruos, într-un proces care, în parte, se autoconsolidează („Din moment ce foarte puțini se deranjează să meargă la vot, eu de ce ar trebui să o fac?"). Putem inversa procesul, parlamentul poate legifera obligativitatea votului și poate impune o amendă pentru aceia care nu merg la vot. Pot supune această

inițiativă unui vot prin referendum. Când votează o chestiune precum obligativitatea votului, cetățenii se confruntă cu o alegere foarte diferită de aceea pe care o au în față atunci când trebuie să voteze în alegerile obișnuite, în care votul nu este obligatoriu. Opțiunile nu sunt „Votez" *versus* „Stau acasă", ci „Toată lumea votează" *versus* „Toată lumea are voie să stea acasă".[1] Dacă mulți dintre aceia care preferă a doua opțiune, în cazul primei alegeri, preferă prima opțiune în cazul celei de-a doua alegeri, atunci aceștia vor *decide colectiv* să facă votul obligatoriu.

Decizia colectivă înseamnă faptul de a face o *alegere a unei politici*. Înainte de implicarea în procesul deciziei colective, fiecare membru are o anumită preferință pentru o politică, generată de preferințele sale fundamentale, împreună cu un set de credințe cauzale cu privire la relațiile dintre scopuri și mijloace. Scopul esențial al deciziei colective este acela de a *agrega* preferințele individuale pentru anumite politici printr-unul dintre cele trei mecanisme care urmează a fi discutate în curând.[2] Agregarea poate determina o *transformare* a preferințelor individuale pentru o anumită politică și poate crea un stimulent pentru indivizi să își *reprezinte greșit* preferințele. Interacțiunea dintre agregare, transformare și reprezentare greșită poate fi deosebit de complexă.

În multe dintre cazurile pe care le voi discuta, un grup mai mic de indivizi ia decizii care sunt obligatorii pentru grupul mai mare. Uneori, aceștia sunt delegați să ia aceste decizii ca reprezentanți sau negociatori pentru grupul mai mare. În acest caz, aceștia pot fi constrânși de cunoașterea faptului că deciziile lor trebuie ratificate de aceia care i-au delegat sau de faptul că nu vor mai fi aleși în cazul în care nu vor reuși să obțină rezultate satisfăcătoare. În alte cazuri, societatea în ansamblu nu are puterea, cu excepția revoluțiilor, de a-i influența pe aceia care iau deciziile care le modelează viețile. Însă chiar și aici putem vorbi despre decizie colectivă la nivelul elitelor. După căderea lui Stalin, a urmat conducerea colectivă a Biroului Politic. Junta chiliană care și-a exercitat puterea din 1973 până în 1980 avea o modalitate internă puternic structurată de a lua decizii colective.

Cele trei mecanisme de agregare la care mă voi referi în cele ce urmează sunt argumentarea, negocierea și votul. Cred că aceasta este o listă exhaustivă. Cu toate că unele decizii colective sunt luate printr-un instrument de randomizare, ca atunci când se aruncă cu moneda pentru a ieși dintr-un balotaj sau când liderii sunt aleși prin tragere la sorți din populația luată în ansamblu, decizia însăși de a adopta o astfel de procedură trebuie să fi rezultat în urma argumentării, negocierii sau votului. Nu voi aborda

---

[1] Cu alte cuvinte, opțiunea „Stau acasă, dar toți ceilalți trebuie să voteze" nu ar fi o opțiune la referendum.

[2] Cu toate că expresia „mecanism de agregare" este asociată de obicei procedurilor de vot, o folosesc aici pentru a mă referi la procesul prin care actorii care pot avea inițial preferințe diferite interacționează pentru a produce o decizie pe care toți o pot accepta ca obligatorie.

interesanta chestiune dacă una dintre aceste trei proceduri este, într-un anumit sens, mai esențială decât celelalte, de exemplu, dacă decizia de a rezolva o anumită chestiune prin vot trebuie luată prin argumentare. Voi investiga, însă, proprietățile fiecărei proceduri, inclusiv elementele lor patologice interesante.

Voi începe prin a da unele exemple pentru aceste trei proceduri. Argumentarea pură se observă (sau se presupune, cel puțin, că ar trebui să fie regula) în cazul proceselor cu juraţi unde se cere unanimitate. Chiar și aici, unii juraţi pot recurge la negocieri tacite prin abilitatea lor mai mare de a-și menține poziția, fiind, cu alte cuvinte, mai puțin nerăbdători să termine cu munca de jurat pentru a se întoarce la viața lor obișnuită.[1] Întrucât timpul este întotdeauna esențial când trebuie luată o decizie și pentru că participanții la proces au de cele mai multe ori preferințe de timp diferite, acest caz ar putea fi chiar cel mai des întâlnit.

Putem ilustra negocierea pură prin jocurile secvențiale de tip „împarte un dolar", în care părțile fac oferte și contraoferte succesive. Rezultatul este determinat de mecanismul de negociere și de puterea de negociere a părților, adică de resursele care le îngăduie să facă amenințări și promisiuni credibile. Acest proces este ilustrat în Figura 20.2.

Votul pur era modul în care Rousseau concepea decizia colectivă. Cetățenii urmează să își formeze preferințele izolat unii de alții pentru a nu fi contaminați de elocvență și demagogie. Întrucât ar fi urmat să își dea votul tot izolat unii de alții, trocul cu voturi avea să fie exclus. În sistemele politice reale, acest ideal nu este niciodată realizat. Am putea ilustra, poate, această situație prin anumite decizii cu miză mică, precum alegerea unor membri ai unei academii științifice a căror funcție ar urma să fie alegerea unor noi membri.

Combinarea argumentului cu votul, fără negociere, poate fi ilustrată prin deciziile de angajare și titularizare în departamentele universităților. Se presupune că acestea sunt guvernate de deliberarea cu privire la meritele candidatului, urmată de un vot. Cu toate că acest ideal nu corespunde întotdeauna realității, uneori corespunde. În departamentele serioase există o normă împotriva favoritismelor, consolidată de o normă împotriva votului acordat fără nicio explicație.

Combinarea argumentării și a negocierii, fără vot, este ilustrată de negocierile colective ale salariilor. Când un sindicat și patronatul decid cum să împartă veniturile unei firme, ar putea părea că doar se negociază. Dacă aruncăm, însă, o privire mai atentă, putem observa că există întotdeauna destul de multă argumentare cu privire la chestiuni de fapt, precum bunăstarea financiară și productivitatea forței de muncă.

---

[1] În primele procese cu juraţi din Anglia, practica înfometării juraților (sau obligarea lor să își plătească mâncarea) până ajungeau la o decizie unanimă se poate să fi conferit unora o putere de negociere mai mare decât altora.

Combinarea negocierii și a votului poate apărea în cazul în care membrii unui sindicat trebuie să ratifice prin vot o convenție salarială negociată de reprezentanții lor. În astfel de situații, rezultatul așteptat al votului servește drept punct de amenințare a negocierii.

Decizia politică, fie cea a unui comitet sau adunare, fie a populației în ansamblul ei, presupune de cele mai multe ori toate trei procedurile.[1] Acest fapt decurge din nevoia de a ajunge la o decizie cât mai rapid posibil. Votul are tendința să apară când o chestiune trebuie decisă urgent, astfel încât participanții nu au timp să delibereze până se atinge unanimitatea. De cele mai multe ori, se întâmplă să existe motivația de a nu se ajunge la unanimitate. Dacă decizia este mai presantă pentru unii participanți decât pentru alții, se ivește posibilitatea pentru negociere întrucât aceia care își permit să aștepte pot cere concesii în schimbul unei decizii timpurii. În comitetele permanente sau în adunări, negocierea apare ca urmare a schimburilor de favoruri întrucât intensitatea preferințelor cu privire la chestiunile care urmează a fi primite în compensație este inegală. Printre alte mecanisme de negociere în legislaturi se numără obstrucționismul și „politica scaunului liber", prin care un anumit grup poate exploata regulile de stabilire a cvorumului pentru a obține ceva ce nu poate obține în alt mod.

În astfel de cazuri, sursele puterii de negociere sunt create în interiorul adunării în sine. În alte cazuri, decidenții pot face apel la resurse care există independent de adunare – bani sau oameni. În 1789, dezbaterile din Adunarea Constituantă franceză au fost suspendate între trupele regelui și mulțimile din Paris, ultimele fiind inițial o armă pe care deputații o putea folosi împotriva celor dintâi, iar, mai târziu, o amenințare pe care unii deputați și-au adresat-o unii altora. În 1989, discuțiile de tip masă rotundă cu caracter cvasiconstituțional sau preconstituțional care au avut loc în Polonia au fost suspendate între amenințarea unei intervenții sovietice și perspectiva înghețării economiei. Dacă un vot nu poate fi cumpărat cu promisiunea altui vot, cum se întâmplă în cazul schimburilor de favoruri, atunci poate fi cumpărat pe bani, de pildă, prin alocarea de fonduri de la partid pentru scopurile unei campanii de realegere în funcție.

În cele ce urmează, se va vedea cum cele trei moduri ale deciziei colective pot fi văzute ca trei pași într-o secvență idealizată, în sensul că fiecare dintre ele decurge natural dintr-un altul, care îl precedă. Cu toate că argumentarea vizează în mod intrinsec atingerea unanimității, în sensul că se bazează pe motive, iar acestea se presupune că sunt valide pentru toți, această țintă este rareori atinsă. Pentru a decide chestiunea, e nevoie de vot. Pentru că votul apare de multe ori între indivizi care au multe chestiuni de decis, este firesc să existe și negociere sub forma schimburilor de favoruri.

---

[1] Chiar și alegerile generale pot lăsa loc negocierilor. Dacă votul este public, votanții și candidații se pot târgui pe prețul voturilor.

Argumentarea

Argumentarea este efortul de a persuada prin prezentarea de motive. De la discursul lui Pericle în fața atenienilor, acest mod de a decide a fost puternic legat de politicile democratice:

> Oamenii noștri publici trebuie să se preocupe, pe lângă chestiunile publice, și de treburile lor private, iar cetățenii noștri de rând, cu toate că sunt ocupați cu diverse îndeletniciri, rămân judecători drepți ai chestiunilor de interes public întrucât, spre deosebire de toate celelalte națiuni, noi nu îi considerăm pe cetățenii care nu iau parte la aceste îndatoriri neambițioși, ci inutili, și suntem capabili să judecăm propuneri chiar dacă nu le-am făcut noi înșine; în loc să privim dezbaterea ca pe o piedică în calea acțiunii, noi ne gândim că este primul pas al oricărei acțiuni înțelepte.

Legătura dintre instituția dezbaterii publice și „acțiunea înțeleaptă" poate fi întrucâtva indirectă. De cele mai multe ori, principalul efect al manifestării în public este acela de a exclude recursul explicit la interesul personal. Într-o dezbatere publică, un vorbitor care spune „Facem acest lucru pentru că-mi convine mie" nu ar convinge pe nimeni și, mai mult, ar fi supus unor sancțiuni informale și ostracizării, ceea ce ar însemna că, în viitor, ar fi încă și mai ineficient. Chiar și aceia care sunt motivați doar de interes sunt constrânși de desfășurarea publică a dezbaterii să își prezinte propunerile de strategii ca fiind motivate de valori mai puțin părtinitoare. Procesul *reprezentării greșite* a preferințelor diferă de acela al *transmutării* (capitolul 4), în același mod în care înșelarea diferă de autoînșelare. Interesul vorbitorului, și nu nevoia de a avea o părere bună despre sine, este acela care îl determină să își reprezinte în mod greșit interesul ca motiv. Interesul său ar putea să îl determine să facă reprezentarea greșită greu de observat, aducând argumente nepărtinitoare cu privire la o strategie oarecum diferită (dar nu prea mult) de aceea care ar coincide perfect cu interesul său personal. S-ar putea întâmpla ca reprezentarea greșită să aibă un efect de recul în cazul în care devine prea evidentă. Prin urmare, *nevoia de a deghiza preferințele fundamentale ale cuiva poate determina o schimbare a preferințelor pentru o anume strategie* prin ceva ce am putea numi „forța civilizatoare a ipocriziei".

În multe societăți, proprietatea a fost utilizată drept criteriu pentru acordarea dreptului la vot. Firește, s-ar putea oferi argumente imparțiale pentru acest principiu. În cadrul Convenției Federale, Madison a argumentat că impunerea unor calificative în funcție de proprietate pentru Senat, iar nu protejarea privilegiaților împotriva poporului, este un mod de a proteja poporul de el însuși. Însă, după cum am mai remarcat, există ceva dubios cu privire la astfel de argumente, care coincid prea bine cu interesul egoist al celor bogați. Poate fi util, mai apoi, să recurgem la alfabetizare ca la un

criteriu imparțial, corelat *foarte puternic, dar nu perfect*, cu proprietatea. În diverse etape ale istoriei americane, alfabetizarea a servit ca element de legitimare pentru scopuri care nu puteau fi recunoscute fățiș, precum dorința de a-i menține pe negri sau pe catolici în afara politicii. Politica americană de imigrare s-a folosit, de asemenea, de alfabetizare ca de un substitut pentru un criteriu care nu putea fi enunțat public. Propunerea de triere a imigranților prin testarea gradului de alfabetizare în limba lor maternă era justificată, de obicei, ca un mod de selectare pe baza meritului individual, o procedură larg acceptată și imparțială. Adevărata motivație a susținătorilor alfabetizării era, însă, înrădăcinată în prejudecată sau în interesul de grup. Anumite grupuri politice au vrut să-i excludă pe imigranți, de obicei analfabeți, din Europa Centrală și de Sud-Est. Laburiștii s-au temut că un aflux de muncitori necalificați ar afecta negativ salariile.

Ar fi greșit, desigur, să ne gândim că argumentarea poate fi întotdeauna redusă la metode mai mult sau mai puțin subtile de promovare a interesului personal. Dacă așa ar sta lucrurile, reprezentarea greșită nu ar avea sens întrucât nimeni nu ar accepta.[1] Dacă vorbitorii sunt motivați de o dorință sinceră de a promova binele public, argumentarea și dezbaterea pot să le schimbe credințele într-un mod care să determine o schimbare a preferințelor politice. Sunt șanse mari ca așa ceva să se producă dacă membrii diferiți ai unui grup au acces la informații diferite, astfel încât ar putea să își îmbunătățească deciziile prin împărțirea în comun a cunoștințelor.[2] În cazul în care organizația este reprezentativă, devine importantă selecția delegaților proveniți din medii diferite. În alegerea reprezentanților pentru o adunare națională, de pildă, această observație este în favoarea votului proporțional cu un prag electoral scăzut sau fără prag.[3] S-ar putea impune reprezentanților dintr-un anumit district electoral să fie membri ai acelei comunități.

Oamenii argumentează, deși poate mai rar, și cu privire la preferințele lor fundamentale, pe care le pot modifica în urma dezbaterii. De multe ori, schimbarea se produce prin descoperirea de asemănări ascunse între diferitele cazuri sau prin expunerea la asemănări superficiale. De pildă, mulți oameni se opun utilizării „organelor prelevate de la cadavre" cu scopul transplantului. Aceștia cred că, dacă familia are obiecții

---

[1] Similar, s-a susținut că, dacă *toate* semnalele animale ar fi înșelătoare, înșelarea nu ar mai avea niciun sens.

[2] Să ne aducem, însă, aminte că sunt șanse mult mai mari pentru îmbunătățire dacă pun la comun datele brute, și nu doar concluziile la care ajung pe baza datelor brute (capitolul 23).

[3] Alte considerații, în special nevoia unei guvernări eficiente, poate favoriza votul majoritar sau votul proporțional cu un prag electoral ridicat. În alegerile pentru adunările constituante, în care guvernarea este o chestiune secundară, se manifestă o tendință de a alege delegați prin vot proporțional.

religioase împotriva acestei practici, atunci sentimentele acesteia ar trebui să fie respectate. Am putea obiecta împotriva acestui punct de vedere aducând în discuție obligativitatea autopsiei în cazul morților dubioase, chiar și atunci când procedura este împotriva credințelor religioase ale familiei. Dacă este nevoie de proceduri invazive pentru a determina cauza morții, se poate argumenta că ar trebui ca acestea să fie acceptabile și pentru a salva vieți. Schimbarea se poate produce și în cazul în care despre un principiu general se crede că vine în contradicție cu intuițiile despre cazuri particulare. O persoană poate accepta obligativitatea utilizării organelor prelevate de la cadavre din motive utilitariste, dar se poate opune ideii că este justificată uciderea unei persoane alese la întâmplare pentru a-i folosi inima, rinichii, plămânii și ficatul cu scopul de a salva viața altor cinci oameni.[1] Prin urmare, un utilitarism inițial simplu poate fi revizuit pentru a ține cont de valori non-consecinționiste (capitolul 4).

Beneficiile argumentării pot fi, însă, subminate prin efectele *vorbitului în public*. Indivizii motivați de bunăstarea publică ar putea, la fel ca alții, să cadă victimă amorului propriu, care le va tăia cheful să recunoască în public că s-au răzgândit. În capitolul 3 am făcut observația că acesta este principalul motiv, identificat de Madison la multă vreme după aceea, pentru care Convenția Federală s-a desfășurat cu ușile închise, iar delegații au jurat să păstreze tăcerea cu privire la lucrurile discutate. Însă argumentul său ar părea să intre în conflict cu un argument tradițional în favoarea accesului publicului la dezbaterile din adunări. Multe decizii legislative au un impact puternic pe termen scurt asupra intereselor legiuitorului. Dacă procesul decizional este ferit de ochii publicului, dezbaterea interesului comun va degenera foarte repede într-o simplă negociere de interese. Acceptarea publicului la dezbateri și modul în care se votează poate limita astfel de comploturi în serviciul interesului personal și poate, ca efect secundar, să promoveze binele public. După cum scria Bentham, „cu cât este mai mare numărul tentațiilor la care este expusă exercitarea puterii publice, cu atât devine mai necesară existența unor motive de a rezista tentației. Însă nu există vreun motiv mai universal și mai constant decât supravegherea publicului." Sau, după cum spune judecătorul american Louis Brandeis, „lumina soarelui este cel mai bun dezinfectant".

Aceste considerații indică o tensiune în procesul de argumentare. Dacă dezbaterile sunt publice, calitatea argumentelor poate avea de suferit. Dacă se țin în spatele ușilor

[1] Utilitariștii au tendința să nege că poate decurge o astfel de consecință. În mod obișnuit, ei susțin că efectele negative ale fricii și incertitudinii care ar fi generate de cunoașterea că ai putea fi ales ca „donator întâmplător" ar fi mai mult decât suficiente pentru a contrabalansa beneficiile unei astfel de practici. Dar cum pot ei ști acest lucru? Suspectez faptul că se folosesc de un raționament invers, dinspre caracterul evident inacceptabil al unei asemenea practici spre existența unor costuri care ar putea să o excludă pe temeiuri utilitariste, și nu un raționament direct, care să pornească de la demonstrarea costurilor și să ajungă la respingerea procedurii.

închise, argumentarea poate degenera în negociere. Tensiunea poate fi, însă, atenuată dacă chestiunile care urmează a fi decise nu îngăduie foarte mult jocul intereselor private. Adunările constituante pot fi mult mai puțin dispuse să ia decizii în interesul propriu decât legislaturile ordinare, nu pentru că delegații sunt motivați imparțial, ci pentru că (sau în măsura în care) interesele lor nu au o influență atât de mare asupra chestiunilor care se decid. De asemenea, după cum am remarcat de mai multe ori, date fiind perspectivele pe termen lung ale construcției constituționale, interesul personal poate, până la un punct, să imite dorința de a promova binele public.

Votul

În cazul în care argumentarea eșuează, poate fi nevoie de vot pentru a se obține consensul asupra unei politici. În votul popular există anumite dimensiuni de variație, printre care se numără: drept electoral, modul în care se votează (secret *versus* la vedere), majoritatea de care este nevoie pentru a se ajunge la o decizie și, în anumite sisteme electorale, cvorumul în cazul unui referendum. În voturile din adunări, dimensiunile principale sunt: cvorumul, dimensiunea majorității și alegerea între votul prin strigarea ordinii de zi și numărarea acelora care ridică mâna (și proceduri asemănătoare, precum „strigatul" sau „ridicatul în picioare"). Votul în secret ca procedură în cadrul adunărilor este rar, dar există și astfel de situații. În Adunarea Constituantă franceză din 1789-1791, președintele adunării a fost ales prin vot secret. Mai recent, Italia și Columbia au practicat votul secret în parlament. De remarcat că trebuie să facem o distincție între votul secret și dezbaterile în spatele ușilor închise, caz în care accesul publicului este interzis. Cel din urmă caz poate fi combinat cu votul public, care îngăduie membrilor adunării să pună pe masă promisiuni credibile în vederea schimburilor de favoruri, care, prin vot secret, ar fi imposibile. Spre deosebire de această situație, dacă dezbaterile sunt deschise publicului, unii auditori pot manifesta o atitudine negativă dacă își văd reprezentanții votând împotriva preferințelor lor într-o anumită chestiune întrucât nu sunt în măsură să observe câștigurile obținute astfel într-o altă chestiune.

În cele ce urmează mă voi limita la votul majoritar. Chiar dacă aceasta nu este o practică universală, decizia de a adopta propuneri printr-o mare majoritate, de exemplu, trei cincimi sau două treimi, pare că trebuie să fie atinsă printr-o majoritate simplă. Adunările constituante care impun adesea majorități calificate pentru amendamentele constituționale ulterioare folosesc aproape invariabil votul prin majoritate simplă în propriile lucrări.[1] Modelul idealizat, în care adunările constituante, în spate-

[1] Constituția sud-africană din 1996 este, în parte, o excepție. Condiția să fie adoptată de o majoritate calificată a fost specificată în Constituția interimară din 1993, adoptată, la rândul ei, prin negociere, și nu prin vot.

le vălului de ignoranță, decid în unanimitate că vor decide prin vot majoritar odată cu ridicarea vălului, are o relevanță foarte mică asupra modului în care se ajunge în realitate la o anumită constituție. Voi ignora chestiunea cvorumului, dar voi face observația că abținerea sau „politica scaunului liber" poate fi folosită de o minoritate pentru a bloca o decizie care ar putea fi adoptată dacă respectiva minoritate ar fi prezentă și ar vota împotriva ei.

Lipsa consensului, care impune necesitatea deciziei prin vot, poate fi determinată de preferințe fundamental divergente, de credințe divergente sau de ambele. Putem lua ca exemplu dezbaterile cu privire la sistemul unicameral sau bicameral purtate în cadrul Adunării Constituante franceze în 1789. La modul foarte general, adunarea era compusă din trei grupuri de dimensiuni relativ egale. Reacționarii voiau să revină la monarhia absolutistă, centrul moderat voia o monarhie constituțională cu prerogative reduse pentru parlament, iar stânga voia o monarhie constituțională cu prerogative sporite pentru parlament. Cu privire la chestiunea bicameralismului, lucrurile stăteau în maniera prezentată simplificat în Tabelul 25.1.

În cele din urmă, bicameralismul a fost înfrânt de alianța dintre reacționari și radicali. Acest fenomen general – acordul politic bazat pe preferințe diferite și credințe diferite care se anulează reciproc – este destul de des întâlnit. Se poate ajunge chiar la unanimitate pe astfel de baze[1], cu toate că diferită de aceea care ar rezulta în urma unei „situații a discursului ideal", în care vorbitorii sunt motivați doar de binele comun și sunt dispuși să asculte argumente.

În versiunea stilizată pe care am supus-o dezbaterii, o majoritate credea că bicameralismul va consolida regimul și o majoritate (diferită) voia să consolideze regimul

TABELUL 25.1

	Preferințe fundamentale	*Credințe*	*Preferințe pentru politici*
Reacționari	Destabilizează regimul	Bicameralismul va consolida regimul	Unicameralism
Moderați	Consolidează regimul	Bicameralismul va consolida regimul	Bicameralism
Radicali	Consolidează regimul	Bicameralismul va destabiliza regimul	Unicameralism

[1] În Adunarea franceză acest rezultat s-a întâmplat în mai 1791, când radicalii, moderații și reacționarii și-au unit forțele pentru a vota legea care făcea ca membrii adunării constituante să fie neeligibili pentru prima legislatură ordinară. Scopul radicalilor era acela de a slăbi legislativul în favoarea clubului iacobinilor, scopul reacționarilor era să îl slăbească în favoarea regelui. Votul a fost unanim întrucât centrul moderat, „beat de dezinteres" (capitolul 5), a votat cu entuziasm pentru a-și nega orice rol în viitorul legislativ.

(vezi Tabelul 25.1). Dacă deciziile colective ar fi fost luate prin agregarea credințelor, ca rezultat al votului majoritar (sincer), apoi prin agregarea preferințelor fundamentale, ca rezultat al votului majoritar (sincer), pentru ca, în final, să se acționeze ținând cont de credințele agregate care ar satisface cel mai bine preferințele agregate, s-ar fi ales *bicameralismul*. (Pentru a împiedica acest rezultat, reacționarii ar fi putut afirma credința falsă că bicameralismul ar destabiliza regimul, formând astfel o majoritate pentru respectiva credință și, prin urmare, o majoritate pentru alegerea unicameralismului.) Din câte știu, această procedură a „dublei agregări" nu este niciodată utilizată în practică. Dată fiind complexitatea agregării directe a preferințelor pentru politici, orice astfel de sistem ar fi, probabil, imposibil de ținut sub control.

În dezbaterea cu privire la bicameralism, credințele votanților erau diferite, la fel și scopurile lor ultime. În alte cazuri, aceștia pot fi asemănători din anumite puncte de vedere și diferiți din altele. Din moment ce preferințele pentru politici sunt, de fapt, observate și agregate, poate fi dificilă clarificarea celor doi factori care intră în componența lor. În mod abstract, putem încerca totuși să determinăm efectele votului majoritar asupra agregării credințelor (dacă asumăm scopuri identice) și asupra agregării preferințelor fundamentale (presupunând credințe identice). Conform lui Tocqueville, democrația (adică votul majoritar, în care dreptul de vot este acordat unei mari majorități) este superior altor sisteme din ambele puncte de vedere. „Ascendentul moral al majorității se bazează în parte pe ideea că o adunare a celor mulți este mai luminată și mai înțeleaptă decât mintea unuia sau că numărul legiuitorilor contează mai mult decât metoda de selecție și pe principiul că interesele celor mulți ar trebui să fie preferate intereselor celor puțini." Oficialii democrați pot „comite erori grave", dar „nu vor adopta niciodată o linie politică ostilă majorității".[1]

Să luăm mai întâi în considerare agregarea credințelor. Există o dezbatere clasică și de durată referitoare la întrebarea dacă, pentru a ajunge la credințele corecte, o procedură mai bună ar fi angajarea unui drept de vot extins sau restrâns, cu alte cuvinte, dacă cei mulți sunt mai înțelepți decât cei puțini. Conform lui Aristotel, avem de-a face cu o chestiune de ponderare a cantității (numărul participanților implicați în procesul politic) și a calității (competența participanților):

> Calitatea poate exista într-una dintre clasele care compun statul, iar cantitatea în cealaltă. De exemplu, cei de condiție inferioară pot fi mai numeroși decât nobilii, iar săracii mai numeroși decât bogații, însă clasa mai numeroasă poate să nu depășească din punctul de vedere al calității atât de mult cât rămâne în urmă

[1] Acesta nu se întreabă și dacă predispoziția de a face greșeli grave ocazional nu ar putea fi chiar mai serioasă decât părtinirea sistematică a unui regim nedemocratic. Cu toate că Tocqueville susține că Statele Unite își permit să facă greșeli datorită situației sale geografice favorabile, acest lucru poate să nu fie adevărat despre alte state.

din cel al calității. Prin urmare, cei doi factori trebuie judecați prin raportare unul la celălalt. Acolo unde mulțimea săracilor depășește în proporția stabilită [astfel încât să compenseze calitatea lor inferioară], este firesc să existe o democrație.

În limbaj modern, putem prezenta chestiunea în termenii „teoremei juraților" a lui Condorcet. Să presupunem că jurații își exprimă credințele (independente) cu privire la acuzațiile procurorului despre faptele inculpatului și că fiecare dintre aceștia are o șansă mai mare de 50% de a avea dreptate. Condorcet a arătat că, dacă jurații decid prin vot majoritar, șansele de a ajunge la rezultatul corect crește odată cu mărimea juriului[1] și converge spre certitudine dacă juriul este indefinit de mare. De asemenea, pentru o anumită dimensiune a juriului, șansa ca majoritatea să aibă dreptate crește când crește șansa fiecărui jurat de a avea dreptate.[2] Prin urmare, după cum sugerează Aristotel, rezultatul poate fi îmbunătățit crescând numărul de jurați sau îmbunătățind pregătirea acestora.[3]

Trecând peste observațiile lui Aristotel, putem observa că pregătirea poate fi mai degrabă o funcție directă a numărului decât a poziției socioeconomice. În limbajul științelor sociale, competența votanților ar putea fi „endogenă" sistemului, și nu „exogenă". Să presupunem că trebuie să alegem între două sisteme politice, oligarhia și democrația, ambele ajungând la decizii prin vot majoritar, dar având dimensiuni diferite ale populației cu drept de vot. Într-o democrație, votanții vor decide în mod rațional să își păstreze ignoranța întrucât impactul fiecăruia asupra rezultatul este foarte mic.[4] Într-o oligarhie, votanții vor investi mai mult în culegerea de informații întrucât fiecare dintre ei va avea un impact mai mare.

---

1 Presupunând, contrar celor enunțate în paragraful următor, că șansa ca fiecare votant să aibă dreptate nu este influențată de creșterea numărului de votanți.

2 Șansa ca majoritatea să aibă dreptate crește și dacă se impune condiția unei majorități calificate, de exemplu, de 60%. Însă, în acel caz, ar putea exista un „juriu în impas", în care nu se atinge majoritatea nici pentru vinovăție, nici pentru nevinovăție.

3 Ne-am putea asigura că sunt satisfăcute condițiile pentru teorema lui Condorcet crescând șansele ca credințele votanților să fie cu adevărat independente unele de altele. Din această perspectivă, putem înțelege propunerea lui Rousseau de a interzice discuțiile înaintea deliberărilor. În același timp, dacă deliberarea îmbunătățește calitatea credințelor, nu se poate obiecta că acestea devin astfel mai puțin independente unele de altele. Condițiile teoremei sunt suficiente pentru ca votul majoritar să producă un rezultat bun, dar nu sunt și necesare.

4 Din moment ce decizia de a vota este în sine irațională (capitolul 12), ne-am putea întreba dacă nu cumva cetățenii ar putea investi în mod irațional în culegerea de informații despre chestiunea în cauză. În contextul de față, însă, problema relevantă este dacă cetățenii ar investi mai mult în cazul în care dimensiunea populației cu drept de vot este restrânsă decât în caz contrar, așa cum, în cazul unei curse apropiate în alegeri, prezența la vot ar putea fi mai mare.

Bentham a observat că acest argument se aplică și votului într-o adunare: „Cu cât este mai mare numărul de votanți, cu atât este mai mică ponderea și valoarea fiecărui vot, cu atât este mai mic prețul său în ochii aceluia care votează și cu atât este votantul mai puțin stimulat în a se asigura că votul său corespunde scopului adevărat și chiar să exercite dreptul de vot." Răspunzând argumentului că o adunare (acesta se gândea la Adunarea Constituantă Franceză din 1789) ar trebui să fie mai numeroasă, din moment ce „probabilitatea înțelepciunii crește odată cu numărul membrilor", acesta scria că „diminuarea pe care aceeași cauză o determină în puterea motivației de a face uz de respectiva înțelepciune contrabalansează acest avantaj". În acest troc între calitate și cantitate va exista o dimensiune a electoratului care maximizează probabilitatea ca votul majoritar să producă credința corectă.[1]

Mai departe, să luăm în considerare agregarea preferințelor prin vot. Două probleme esențiale și puternic înrudite se desprind de aici, și anume *reprezentarea greșită* a preferințelor și *indeterminarea* rezultatului. În ce o privește pe prima, oamenii pot fi stimulați să voteze pentru alte propuneri sau pentru alți candidați decât acelea pe care le-ar prefera să fie adoptate sau decât aceia pe care i-ar prefera aleși. Alegerea unui sistem de vot la vedere în locul unui sistem secret poate determina un astfel de fenomen. În Atena clasică, cele mai multe decizii ale adunării se făceau prin ridicarea mâinii, cu rezultatul că unii cetățeni se puteau teme să voteze conform conștiinței lor. Astfel, Tucidide scrie că, „privind entuziasmul majorității [pentru expediția în Sicilia], aceia cărora nu le era pe plac s-au temut să pară lipsiți de patriotism ridicându-și mâinile împotriva ei". (Spre deosebire de aceasta, decizia în tribunalele ateniene cu juriați se făcea prin vot secret.) Alegerea votului prin apel nominal, față de alte metode, precum „ridicatul în picioare", poate să-i intimideze pe votanți, de asemenea. În adunările constituante din Paris (1789-1791) și Frankfurt (1848), extremiștii cereau adesea apelul nominal când venea vorba despre chestiuni importante cu amenințarea implicită – și uneori explicită – că îi vor expune pe aceia care votează împotriva propunerilor extremiștilor violenței populare făcând publice listele cu numele lor. Chiar dacă exista o majoritate clară în sistemul „ridicării în picioare", care îngreuna numărarea voturilor, rezultatul putea fi inversat prin apel nominal.[2]

---

[1] În sens abstract, punctul optim ar putea fi la unul dintre capetele extreme – ori un singur individ, ori toate persoanele adulte. Considerând presupoziții rezonabile, cel mai probabil există un „maximum intern". Dacă dimensiunea optimă este mică, am putea alege votanții la întâmplare dintre cetățeni pentru a ne asigura că aceștia nu reprezintă interese de grup. Din această perspectivă, votul ar fi o *funcție*, și nu un *drept*.

[2] Pentru a da un exemplu mai recent, când s-a dezbătut Legea drepturilor civile din 1964 în Camera Reprezentanților din Statele Unite, președintele Comitetului de Reglementare (Howard Smith) a introdus un amendament pe care l-a crezut ucigător: a propus ca legea să interzică

Reprezentarea greșită poate apărea și în cazul votului secret. În toate sistemele de vot pot apărea situații în care un alegător, votând pentru o altă soluție decât prima în ordinea preferințelor, poate produce un rezultat pe care îl preferă aceluia care s-ar fi produs dacă ar fi votat sincer.[1] (Ar putea apărea o excepție dacă propunerile sau candidații s-ar alege printr-un instrument de randomizare, în care probabilitatea ca o alternativă să fie aleasă este egală cu proporția votanților care o preferă. Într-un asemenea caz, nu ar apărea problema „votului irosit". Dezavantajele sistemului sunt evidente și constituie o explicație pentru faptul că nu a fost ales niciodată.) Dorința de a-ți vedea primele opțiuni câștigătoare în ordinea preferințelor cu o marjă care nu e prea mare poate determina un vot împotrivă. În capitolul 18 am menționat, de pildă, modul în care socialiștii ar putea vota cu comuniștii pentru a-și determina partidul să adopte o politică mai de stânga. Dacă este sigur că opțiunea cea mai preferată nu va fi aleasă, poți vota pentru cea mai bună alternativă dintre acelea care au șanse de câștig. Unele sisteme de vot creează un stimulent de a clasifica un candidat preferat sau o propunere preferată de alți votanți într-un mod mai puțin favorabil decât ar dicta preferințele reale ale votantului sau de a introduce alternative noi cu simplul scop de a crește șansele alegerii alternativei preferate.

În unele cazuri, rezultatul votului majoritar este nedeterminat. Să presupunem că există trei blocuri de dimensiuni relativ egale în adunarea municipală, care reprezintă comunitatea de afaceri, muncitorii din industrie și, respectiv, liber-profesioniștii. Adunarea trebuie să decidă între construirea unei piscine acoperite, alocarea unor subvenții pentru orchestra simfonică din localitate sau construirea unui teren de golf. Să presupunem, în conformitate cu stereotipurile aferente fiecărui grup, că (după lungi dezbateri) opțiunile lor sunt ordonate în maniera prezentată în Tabelul 25.2.

---

discriminarea pe criterii de sex, rasă și origine etnică. Conducerea Camerei, împărtășind convingerea lui Smith că amendamentul ar putea ucide propunerea legislativă, a încercat să îi convingă pe democrații liberali să voteze împotriva propunerii și, conform anumitor relatări, mulți dintre acești membri ar fi fost dispuși să o facă dacă un astfel de vot s-ar fi putut face în secret. Însă galeriile erau pline de grupuri de interese ale femeilor, care erau de față pentru a observa comportamentul cu privire la voturile neînregistrate; în aceste circumstanțe, membrii liberali nu s-au arătat dispuși să fie de acord cu conducerea Camerei și să voteze amendamentul lui Smith. Din câte se pare, amendamentul lui Smith nu a dus la respingerea propunerii legislative, ci la extinderea categoriilor protejate de lege, astfel încât să se includă și criteriul de gen.

[1] Mai tehnic, putem spune că, în sistemele de vot nedictatoriale și neprobabiliste, exprimarea preferințelor reale nu este, în general, strategia dominantă. Se pot, însă, concepe sisteme prin care sinceritatea devine un echilibru, în sensul că devine răspunsul cel mai bun la sinceritatea celorlalți. Detaliile sunt complicate și nu reflectă trăsăturile sistemelor de vot întâlnite în lumea reală.

În cazul în care alternativele sunt opuse celorlalte în voturi-pereche, există o majoritate compusă din oameni de afaceri și muncitori, care preferă terenul de golf orchestrei, o majoritate de oameni de afaceri și liber profesioniști, care preferă orchestra piscinei, și o majoritate de liber-profesioniști și muncitori, care preferă piscina terenului de golf. Prin urmare, „preferințele sociale" sunt *intranzitive* sau *circulare*. În cazul alegerilor individuale, tranzitivitatea este o condiție a raționalității (capitolul 11). În contextul de față, problema nu este atât de mult una de raționalitate, cât de determinare.

Dacă consiliul local nu are la dispoziție pentru luarea unei decizii decât Tabelul 25.2, este greu de văzut cum anume va ajunge la o decizie. Din moment ce se votează (pentru că nu se poate ajunge la un consens în consiliu), continuarea dezbaterii este inutilă. Dacă s-ar putea măsura intensitatea cu care diversele grupuri preferă una dintre opțiuni celeilalte sau gradul în care opțiunile satisfac nevoi obiective, s-ar putea argumenta că o anumită opțiune este superioară celorlalte fără dubii. Însă nu există vreo procedură care să permită o comparație între gradele de intensitate ale preferințelor sau ale satisfacerii nevoilor de la individ la individ.[1] De exemplu, dacă le-am cere să răspundă cât de mult valorizează opțiunile punându-i să le clasifice pe o scară de la 0 la 10, ar fi inutil. În primul rând, nu putem ști dacă un anumit punctaj (să zicem, 7) înseamnă același lucru pentru membrii celor trei grupuri. În al doilea rând, dacă îi rugăm să ordoneze opțiunile, le-am crea un stimulent să își reprezinte greșit intensitatea preferințelor, de exemplu, atribuind 10 puncte opțiunii preferate și 0 celorlalte.

În practică, nu este clar cât de importantă este problema „preferințelor sociale ciclice". Aceasta nu poate apărea dacă preferințele individuale au „un singur vârf", adică opțiunile pot fi ordonate de la „cea mai de sus" la „cea mai de jos", astfel încât preferințele fiecărui individ să crească treptat spre politica preferată și să descrească treptat pe măsură ce ne îndepărtăm de aceasta. În multe cazuri, aceasta este o proprietate

TABELUL 25.2

	Oameni de afaceri	*Muncitori*	*Liber-profesioniști*
Teren de golf	1	2	3
Orchestră	2	3	1
Piscină	3	1	2

[1] Acesta este motivul pentru care utilitarismul clasic ce recomandă maximizarea sumei totale a utilităților individuale este dificil de aplicat în realitate. Această problemă nu ar trebui să ne facă să ne gândim că este întotdeauna lipsit de sens să comparăm gradele de bunăstare de la o persoană la alta. Atât instituțiile, cât și indivizii fac astfel de comparații în mod obișnuit, cu un grad mare de încredere. Însă, întrucât comparațiile se bazează pe intuiții, și nu pe măsurători, vor exista multe cazuri în care acestea vor eșua în a produce rezultate care să nu fie ambigue.

rezonabilă a preferințelor. Dacă planul de taxare preferat de un individ este cel de 20%, acesta poate prefera 19% în loc de 18% și 21% în loc de 22%. Mai mult, nu există cazuri în care membrii unei adunări să își ridice mâinile în aer declarând că, din moment ce nu există „voință populară", nu se va lua nicio decizie. De fapt, dacă una dintre opțiuni este să nu se schimbe nimic, ideea este chiar incoerentă. O decizie se va lua întotdeauna, fie în mod implicit (păstrând lucrurile așa cum sunt), fie prin adoptarea unei proceduri tradiționale de vot, fie prin manipularea agendei.

Însă faptul că se ajunge la o decizie nu înseamnă că ea reprezintă voința populară sau „generală" într-un mod nearbitrar. Pentru o constelație de preferințe (sincere), precum cele din Tabelul 25.2, ideea în sine a unei voințe generale este lipsită de sens. De câte ori întâlnim astfel de constelații de preferințe? Cercetătorii din științele politice ne pun la dispoziție nenumărate exemple. Alții au argumentat că așa-zisele exemple nu sunt descrise corect și că, la o examinare mai atentă, ar trebui să respingem ce se susține despre majoritățile ciclice. Voi descrie, în cele ce urmează, două cazuri care par a fi instanțe autentice de preferințe ciclice.

În 8 octombrie 1992, parlamentul norvegian a decis ca viitorul aeroport din zona Oslo să fie localizat în Gardermoen (voi simboliza această alternativă prin litera G). Alte opțiuni erau Hobol (alternativa H) și o soluție care presupunea combinarea opțiunii Gardermoen cu aeroportul Fornebu, deja existent (alternativa D). Opțiunile aveau să fie luate în considerare succesiv, în raport cu starea de fapt, și nu evaluate prin raportarea lor una la alta. Odată ce o anumită opțiune urma să primească o majoritate de voturi, avea să fie adoptată.[1] Cu toate că acest tip de vot în serie era sistemul tradițional de vot în parlament, erau posibile și alte sisteme, ca, de exemplu, compararea opțiunilor în perechi, până când una avea să iasă câștigătoare. Votând în serie, ordinea în care sunt prezentate opțiunile poate fi decisivă, după cum vom vedea imediat.

Preferințele *exprimate* de partide, care coincideau cu unele excepții neimportante cu votul deputaților, erau cele de mai jos:

Partidul Laburist (63 de deputați): G > D > H
O coaliție a Partidului Socialist de stânga, a creștin-democraților și a Partidului Agrar (42 de deputați): D > H > G
Partidul Conservator (37 de deputați): H > D >G
Un deputat independent: G > H > D

Presupunând că acestea sunt preferințele *sincere*, atunci preferințele sociale sunt ciclice: D învinge H cu 105 la 60, H învinge G cu 101 la 64 și G învinge D cu 101

[1] Din moment ce toate erau preferate în locul situației de fapt, cu siguranță una dintre ele urma să fie adoptată.

la 64. Înainte de supunerea la vot, parlamentul a votat ordinea în care ar trebui luate în considerare alternativele. Laburiștii au propus G-D-H, în timp ce președintele parlamentului a propus D-H-G. Votul între cele două opțiuni a fost câștigat de laburiști. Dacă ar fi ieșit câștigătoare propunerea președintelui, laburiștii probabil că ar fi votat pentru D din moment ce, altminteri, eșecul de a strânge o majoritate pentru D ar fi dus la adoptarea opțiunii cel mai puțin preferate de laburiști, anume H. Pe ordinea astfel adoptată, Partidul Conservator se afla în fața unei dileme asemănătoare. În cele din urmă, conservatorii au votat pentru G pentru că, dacă ar fi votat împotriva acestei opțiuni, atunci ar fi câștigat propunerea cel mai puțin preferată de ei, anume D. Cu toate că este posibil, în abstract, ca laburiștii să fi fost nesinceri în susținerea opțiunii D ca a doua preferată și că au procedat astfel doar pentru a-i face pe conservatori să creadă că, dacă votează împotriva opțiunii G, s-ar adopta D, nu există dovezi pe baza cărora să putem conchide acest lucru. Dacă lucrurile ar fi stat astfel, preferințele sociale nu ar fi fost ciclice, din moment ce H ar învinge atât D, cât și G.

Al doilea exemplu exclude aproape complet ideea că preferințele ciclice ar fi doar un simplu artefact al reprezentării greșite a preferințelor. Cazul este acela al contextului în care era nevoie de o decizie asupra ordinii în care urma să se facă demobilizarea din armata americană după cel de-al Doilea Război Mondial. Părăsirea timpurie a armatei era un beneficiu rar întâlnit, care trebuia alocat în mod echitabil. Pentru a determina criteriile, armata a organizat sondaje pe scară largă în rândul bărbaților recrutați. Într-un sondaj în care criteriile erau opuse unul altuia în perechi comparative, ordonările arătau o incoerența colectivă. Astfel, 55% credeau că un bărbat căsătorit cu doi copii care nu văzuse încă frontul ar trebui eliberat înaintea unui bărbat necăsătorit cu două campanii armate la activ, 52% considerau că optsprezece luni de serviciu în afara granițelor sunt mai importante decât doi copii, iar 60% credeau că două campanii armate valorează mai mult decât optsprezece luni de serviciu în afara granițelor. Este foarte probabil ca respondenții să-și fi reprezentat greșit preferințele.[1]

[1] Autorii studiului din care am luat acest exemplu scriu că „nu ar trebui așteptat un grad mare de coerența internă pentru astfel de alegeri ipotetice foarte complexe", sugerând că problema era una a ordonării individuale a preferințelor. Dacă majoritățile adunate ar fi dus la peste 200%, poate că această sugestie ar fi fost justificată. Din moment ce se adună la doar 167%, este foarte probabil ca ordonările să fi fost coerente individual și să dea naștere unei intranzitivități colective. Studiul a fost publicat în 1949, cu doi ani înainte de lucrarea deschizătoare de drumuri a lui Kenneth Arrow cu privire la agregarea preferințelor și incoerența la care este vulnerabilă.

Negocierea

Negocierea este procesul prin care se ajunge la un acord prin amenințări și promisiuni credibile. Un soț poate amenința cu procesul asupra custodiei unice a copilului în cazul în care partenera nu este de acord să împartă custodia. În negocierile pentru salarii, muncitorii pot amenința cu greva, pot să nu depășească norma sau să refuze ore suplimentare, în timp ce angajatorii pot amenința cu greva patronală sau cu închiderea fabricii. Patronatul unei firme poate amenința să concedieze un angajat dacă acesta nu muncește mai cu sârg. O țară poate amenința să invadeze altă țară dacă acesta nu face concesii teritoriale. Într-o adunare constituantă, un delegat dintr-o unitate teritorială poate amenința să părăsească adunarea dacă aceasta nu adoptă un mod de reprezentare avantajos pentru respectiva unitate teritorială. Senatorii americani pot amenința cu obstrucționarea votării unei legi dacă președintele nu retrage o anumită numire în funcție. Congresul poate amenința că refuză să voteze bugetul dacă președintele se folosește de dreptul de veto pentru a se sustrage de sub incidența unei anumite legi.

Cât privește promisiunile, un membru al unui grup care decide prin vot poate promite să voteze pentru o propunere mai importantă pentru unul dintre colegi, cu condiția ca respectivul coleg să voteze ulterior pentru o chestiune importantă pentru el (schimbul de favoruri). Vânzătorul unei case poate promite să nu înceapă renegocierile în cazul în care un cumpărător e dispus să plătească prețul pe care îl cere. Similar, un răpitor poate promite să dea drumul victimei odată ce a fost plătită răscumpărarea, în loc să rețină victima pentru a impune noi cereri. De asemenea, un guvern poate promite să elibereze din închisoare un terorist odată ce membrii grupării teroriste au dat drumul victimei răpite. Victima unei răpiri, odată eliberată, poate promite să nu descrie modul în care arată răpitorii la poliție. O persoană într-o situație de Dilemă a Prizonierului poate promite să coopereze dacă și cealaltă cooperează.

Rezultatul negocierii depinde de *credibilitatea* amenințărilor și a promisiunilor. O amenințare este credibilă dacă agentul care o face se așteaptă să ducă la îndeplinire amenințarea (nu contează motivul) în cazul în care celălalt nu se supune. „Dă-mi banii sau îți iau viața" este o amenințare mult mai credibilă decât „Dă-mi banii sau îmi iau viața". O promisiune este credibilă dacă se așteaptă ca agentul să își țină promisiunea (nu contează motivul) după ce cealaltă persoană și-a respectat angajamentele. „Voi desena ilustrațiile după ce ai terminat cu textul" este o promisiune credibilă din moment ce persoana care promite va avea un stimulent pentru a-și ține promisiunea. În Jocul Încrederii discutat în capitolele 5 și 15, Administratorul nu are niciun stimulent de ordin material pentru a-și respecta promisiunea de a returna jumătate din câștigurile create de transferul Investitorului.

Credibilitatea *amenințărilor* depinde atât de factori obiectivi, cât și de factori subiectivi. Factorii obiectivi se împart în opțiuni externe și opțiuni interne. O opțiune externă este ceva obținut de o anumită parte la negociere, dacă negocierile ar fi sistate irevocabil. În negocierile de salarii, opțiunea externă pentru muncitori ar fi salariul pe care l-ar putea obține în altă firmă sau nivelul indemnizațiilor de șomaj. În negocierea custodiei copilului, opțiunile externe sunt definite prin rezultatul așteptat al procesului. La Convenția Federală din Philadelphia, opțiunea externă era avantajul pe care îl putea obține un stat dacă se retrăgea din uniune (și, posibil, intrând într-o alianță cu altă națiune). O amenințare de retragere de la masa negocierilor este credibilă doar dacă partea poate demonstra că i-ar fi mai bine în afara relației decât i-ar fi dacă ar accepta oferta celeilalte părți.

O opțiune internă este ceva disponibil părților în timpul procesului de negociere în sine și care este hotărâtor pentru cât de mult pot rezista în negociere. Pentru muncitori, opțiunea internă poate fi hotărâtă de dimensiunea fondului de grevă. Pentru părinți, poate fi alocarea custodiei temporare unuia dintre ei. În Philadelphia, opțiunea internă era determinată de dorința delegaților de a se întoarce la chestiunile presante de zi cu zi din statele de unde proveneau. Opțiunile interne afectează credibilitatea amenințărilor întrucât hotărăsc dacă o parte este stimulată să le aducă la îndeplinire. O amenințare cu grevă a muncitorilor poate să nu fie credibilă dacă (firma știe că) aceștia nu au la dispoziție un fond de grevă și cei mai mulți dintre ei sunt căsătoriți și au de plătit rate la casă.

În general, amenințarea de a produce anumite daune unei alte părți nu este credibilă dacă presupune în același timp daune substanțiale aduse propriei persoane. Membrii Congresului nu pot amenința în mod credibil să refuze votarea bugetului dacă președintele știe că această măsură ar aduce prejudicii serioase atât reputației acestora, cât și capacității lor de a-și realiza proiectele. Un membru (Lanjuinais) al Adunării Constituante franceze a susținut în 1789 că viitorul legislativ nu va putea să își utilizeze controlul asupra veniturilor pentru a preveni un veto din partea regelui. Din moment ce „întreruperea plății taxelor" pentru acest scop este ca și cum „ți-ai tăia gâtul pentru a opri o sângerare la picior", amenințarea de a face așa ceva nu ar fi credibilă.

Din punct de vedere subiectiv, amenințările sunt mult mai credibile dacă agentul are (și celelalte părți cred că are) un orizont larg de timp. Negociatorii răbdători au un avantaj competitiv, din moment ce oponenții lor pot fi dispuși să accepte un schimb între *dimensiunea* unei concesii și *momentul în timp* la care concesia este făcută.[1] În negocierile de la Paris dintre Statele Unite și Vietnamul de Nord, acesta din urmă a

[1] Această formulare poate duce întrucâtva în eroare. Într-un joc de negociere de tipul celui din Figura 20.2, acordul va fi atins imediat pe baza inducției inverse, în care rata deprecierii timpului a fiecărei părți joacă un rol important. Negocierea este „virtuală", nu reală. Însă

făcut o foarte bună mișcare de deschidere când delegația sa a încheiat *un contract pe doi ani* pentru o casă, dând astfel un semnal că nu se grăbește nicăieri. Neutralitatea față de risc sau un grad redus de aversiune față de risc pot, de asemenea, să confere un avantaj competitiv unei părți aflate în negociere. Un părinte cu aversiune față de risc și care are șanse bune de a câștiga custodia deplină a copilului dacă merge la proces ar putea, cu toate acestea, să împartă custodia (capitolul 11). Un alt factor subiectiv care poate conferi credibilitate amenințărilor este atribuirea unei valori modeste ofertei celeilalte părți în comparație cu ceea ce ar trebui cedat. În negocierile de după divorț, un părinte poate fi în măsură să forțeze o înțelegere financiară favorabilă pentru că celălalt părinte ține foarte mult la custodia copiilor.

Orizontul larg de timp poate afecta credibilitatea și în alt mod. Dacă eșuez să duc la îndeplinire o amenințare într-o interacțiune curentă pentru că presupune costuri prea mari, abilitatea de a face amenințări în interacțiuni viitoare ar putea să sufere. Ideea ar putea fi înțeleasă, probabil, mult mai bine dacă inversăm termenii: aș putea duce la îndeplinire o amenințare într-o interacțiune curentă, chiar dacă nu am vreun stimulent să o fac, doar pentru a-mi construi reputația că sunt cineva care ia în serios o amenințare. Dacă oponentul înțelege logica acestui argument, e posibil să nu fiu nevoit să duc la îndeplinire vreo amenințare. Nu vorbim aici despre consolidarea unei reputații în sens literal, ci doar într-un sens „virtual". Cel mai adesea, s-ar putea să fie nevoie să duc la îndeplinire o serie de amenințări „împotriva interesului meu" pentru a da lumii un semnal că sunt acel gen de om. În comentariile despre Jocul Lanțului de Magazine din capitolul 20 m-am referit la ambele posibilități.

Credibilitatea *promisiunilor* depinde și ea atât de factori subiectivi, cât și de factori obiectivi, dar într-un mod oarecum diferit. Credibilitatea unei promisiuni depinde în mod crucial de capacitatea agentului de a putea duce la îndeplinire promisiunea. Să ne gândim, de pildă, la încercarea eșuată a schimburilor de favoruri din Adunarea Constituantă franceză în toamna lui 1789. În trei întâlniri între liderul moderaților, Mounier, și extremiștii Barnave, Duport și Alexandre Lameth, ultimii trei au făcut următoarea propunere: îi vor oferi lui Mounier atât un veto absolut pentru rege, cât și bicameralism dacă va accepta ca regele să renunțe la privilegiul de a dizolva Adunarea, iar Camera Superioară va avea doar un drept de veto provizoriu, și dacă vor exista convenții periodice pentru revizuirea Constituției. Mounier a refuzat fără să stea pe gânduri, probabil pentru că nu credea că cei trei vor fi în măsură să își respecte promisiunea, din moment ce Adunarea nu era compusă din partide în sensul modern al grupurilor disciplinate, care pot fi determinate să voteze în bloc o anumită chestiune.

---

rămâne în picioare ideea că partea mai nerăbdătoare va primi mai puțin, dat fiind că mai important pentru ea este să obțină ceva cât mai repede.

Pentru un alt exemplu, ne putem gândi la promisiunile de imunitate politică făcute foștilor lideri în cazul tranziției la democrație. Promisiuni în acest sens au fost făcute, acceptate și încălcate în Argentina în 1983, în Uruguay în 1984 și în Polonia și Ungaria în 1989. (În America Latină, amenințările cu lovituri de stat militare au forțat ulterior acordul.) Privind în urmă, generalii și liderii partidelor politice ar fi trebuit să înțeleagă că asemenea promisiuni nu sunt credibile întrucât liderii care urmau să preia puterea nu puteau garanta că tribunalele și parlamentele le vor respecta promisiunile. În Polonia, negociatorii opoziției în Negocierile de la Masa Rotundă, care făceau parte din aripa stângă a Solidarității, au susținut că *pacta sunt servanda*, „promisiunile trebuie respectate". Când aripa dreaptă a ajuns la putere angajamentul a fost ignorat.

Printre condițiile subiective ale credibilității se numără, din nou, un orizont larg de timp. Dacă agentul are o rată redusă a deprecierii timpului și știe că va mai interacționa cu persoana căreia îi face o promisiune (sau cu alte persoane care îi pot observa comportamentul), atunci are un stimulent în a-și păstra promisiunea. Schimbul de favoruri la nivelul parlamentului, de pildă, poate fi menținut prin astfel de anticipări. (În adunările constituante, care se întrunesc o singură dată, stimulentul de a-ți ține promisiunea că vei întoarce favoarea ar trebui să fie mai mic și, prin urmare, ar trebui să existe mai puține astfel de promisiuni.) Un agent poate fi în măsură să creeze un mijloc de constrângere pentru sine, astfel încât să nu își încalce o promisiune. Cu toate că, de obicei, răpitorii nu au vreun motiv să creadă promisiunea victimei că, odată eliberată, nu le va dezvălui identitatea, aceasta poate facă promisiunea credibilă dându-le răpitorilor o informație prejudiciabilă și posibil de verificat despre sine, pe care *aceștia* ar putea să o folosească în cazul în care sunt prinși de poliție.

Întrucât condițiile subiective sau stările mentale care modelează rezultatul unei negocieri nu pot fi observate direct, negociatorii au interesul de a le reprezenta greșit, prin comportamente verbale sau nonverbale. Din câte știm, negociatorii Vietnamului de Nord au închiriat o casă pe termen lung doar *pentru a da impresia* că sunt foarte răbdători. În cazul schimburilor de favoruri, fiecare parte poate exagera importanța lucrului care i se cere pentru a forța o concesie mai mare din partea celeilalte părți. Când muncitorii susțin că acordă o mare importanță măsurilor costisitoare de siguranță la locul de muncă, aceasta poate fi doar o strategie prin care se poate justifica un bonus substanțial la salariu în locul respectivelor măsuri. În multe cazuri, încercările de a înșela pot fi prea transparente pentru a funcționa. Dacă un părinte care divorțează susține că este foarte preocupat pentru obținerea custodiei copiilor doar pentru a obține un aranjament financiar mai favorabil, celălalt părinte poate fi în măsură să aducă dovezi care atestă lipsa de interes pentru copil a primului înainte ca mariajul să înceapă să se destrame sau să aducă în discuție recenta acceptare a unei funcții care presupune numeroase deplasări. Un părinte prevăzător poate anticipa, însă, această

problemă și poate pune bazele prin care poate să susțină că îi pasă de copii înainte ca celălalt părinte să înțeleagă că mariajul începe să se destrame.

La fel ca părțile angajate în argumentare, negociatorii pot fi stimulați să își reprezinte interesele greșit din principiu. Raționamentul din spatele reprezentării greșite este diferit, totuși. În cazul argumentării, părțile vor să prevină oprobriul la care ar fi supuse dacă și-ar baza propunerile doar pe interes. În negociere nu există niciun fel de oprobriu dacă îți exprimi doar interesul. Este normal ca firmele și angajații să fie preocupați de profituri și salarii, și nu de binele comun. Negociatorii pot, însă, câștiga un avantaj strategic dacă își prezintă doleanțele în termenii unor principii. Aceștia pot susține că fac o concesie mult mai mare decât oponentul dacă renunță la o cerință susținută de un principiu, așteptând astfel concesii mai mari din partea celuilalt, decât dacă în joc ar fi doar interesul personal. Însă, dacă fiecare parte angajează o astfel de tactică, negocierea poate eșua.

Rezumat

Adunând la un loc diversele argumente prezentate în acest capitol, procesul deciziei colective poate fi reprezentat ca în Figura 25.1. Argumentul central este, poate, acela că fiecare mecanism al deciziei colective – argumentarea, votul și negocierea – creează un stimulent de a reprezenta greșit un anumit aspect al preferințelor. Cu alte cuvinte, *un mecanism de agregare funcționează astfel încât elementele introduse se conformează mecanismului în sine*. Preferințele politice exprimate sunt influențate atât de preferințele politice reale, cât și de mecanismul care agregă preferințele politice

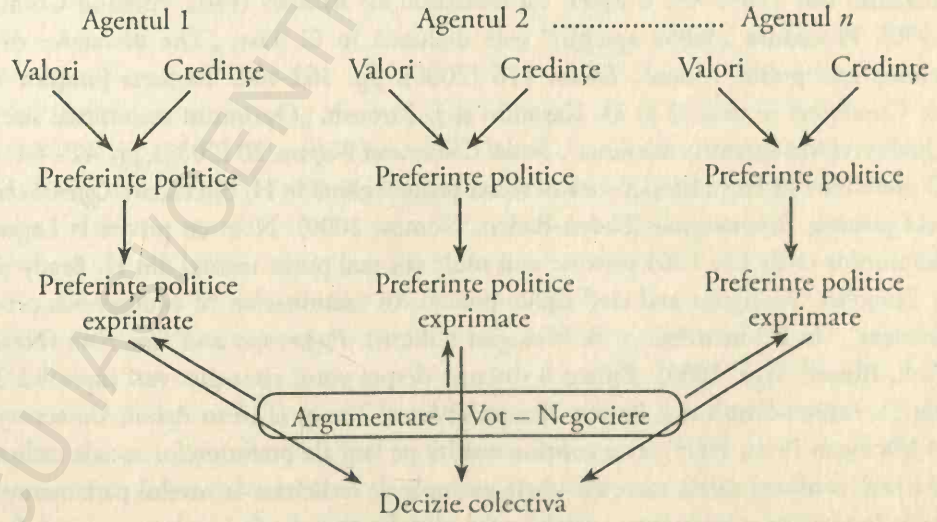

FIGURA 25.1

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

exprimate. Impactul reprezentării greșite asupra bunăstării este ambiguu. În virtutea forței civilizatoare a ipocriziei, efectele pot fi benefice din punct de vedere social. În alte cazuri, utilizarea generalizată a acestei tactici poate duce la o situație de tip Dilema Prizonierului în care toată lumea pierde.

◈◈◈

Notă bibliografică

Pentru mecanisme care sunt, în anumite privințe, intermediare între acțiunea colectivă și decizia colectivă, vezi E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990). Pentru o trecere în revistă a țărilor în care votul este obligatoriu și pentru sancțiunile care li se aplică acelora care nu votează, vezi http://www. idea.int/vt/compulsory_voting.cfm. O discuție clară, cu toate că uneori excentrică, cu privire la argumentare și vot este cartea lui J. Bentham, *Political Tactics* (Oxford University Press, 1999). Pasajele citate din Bentham (traduse din franceză) se regăsesc în la fel de interesanta *Rights, Representation and Reform* (Oxford University Press, 2002), pp. 35 și 122. Pentru reprezentarea greșită determinată de deliberare, vezi capitolul 5 din cartea mea, *Alchemies of the Mind* (Cambridge University Press, 1999). Pentru un studiu al diferenței dintre legislaturile ordinare și adunările constituante, vezi articolul meu „Constituent legislatures", în R. Bauman și T. Kahana (editori), *Constitution and the Legislature* (Cambridge University Press, 2006). Pentru votul cu privire la bicameralism din 1789, vezi J. Egret, *La révolution des notables* (Paris, Armand Colin, 1950). Procedura „dublei agregări" este discutată în C. List, „The discursive dilemma and public reason", *Ethics*, 116 (2006), pp. 362-402. Teorema juraților a lui Condorcet se discută în D. Karotkin și J. Paroush, „Optimum committee size: Quality-versus-quantity dilemma", *Social Choice and Welfare*, 20 (2003), pp. 429-441. O prezentare pe larg a istoriei votului secret poate fi găsită în H. Buchstein, *Öffentliche und geheime Stimmangabe* (Baden-Baden, Nomos, 2000). Nota cu privire la Legea drepturilor civile din 1964 provine mai mult sau mai puțin textual din H. Brady și J. Ferejohn, „Congress and civil rights policy: An examination of endogenous preferences", în I. Katznelson și B. Weingast (editori), *Preferences and Situations* (New York, Russell Sage, 2005). Pentru o discuție despre votul strategic, vezi capitolul 2 din D. Austen-Smith și J. Banks, *Positive Political Theory II* (Ann Arbor, University of Michigan Press, 2003), care conține analize pe larg ale preferințelor sociale ciclice și o teză conform căreia toate așa-zisele exemple de ciclicitate la nivelul parlamentelor se bazează pe o interpretare greșită a faptelor. Exemplul referitor la aeroportul din Oslo provine din A. Hylland, „The Condorcet paradox in theory and practice", în

J. Elster *et al.* (editori), *Understanding Choice, Explaining Behavior: Essays in Honor of Ole-Jørgen Skog* (Oslo Academic Press, 2006). Exemplul referitor la demobilizarea soldaților americani este preluat din S. Stouffer (ed.), *The American Soldier* (Princeton, NJ, Princeton University Press, 1949). Lucrarea esențială pe subiectul negocierii este T. Schelling, *The Strategy of Conflict* (Cambridge, MA, Harvard University Press, 1960). O lucrare clasică despre negocierea în acțiune este H. Raiffa, *The Art and Science of Negotiation* (Cambridge, MA, Harvard University Press, 1982). O introducere de popularizare a teoriei negocierii se poate găsi în cartea mea *The Cement of Society* (Cambridge University Press, 1989).

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Capitolul 26

ORGANIZAȚII ȘI INSTITUȚII

Organizații și instituții

Din rândul organizațiilor fac parte supermarketul din localitate, compania General Electric, Congresul american, Biserica Catolică, Curtea Constituțională a Germaniei, Franța și Națiunile Unite. Exemple de instituții sunt: familia, constituțiile, banii de hârtie, economia de piață și justiția de tranziție. Ne putem referi la organizații și instituții ca instanțe (*tokens*) sau ca tipuri (*types*), adică le putem privi ca instanțe ale unui concept sau drept conceptul în sine. Bursa este o instituție (un tip), Bursa din New York este o organizație (o instanță a respectivului tip). Statul-națiune este o instituție, Franța este o organizație. Însă unele organizații, precum Națiunile Unite, nu sunt instanțe ale vreunei instituții. Inversând termenii, unele instituții, precum economia de piață, nu au instanțe organizaționale. Piețele speciale, însă, precum piața pentru certificate de poluare, pot fi organizații. Cu toate că, în cele ce urmează, mă voi referi mai ales la organizații, le putem înțelege cel mai bine pe fundalul unor instituții.

Organizațiile sunt *actori colectivi*. Ele sunt definite prin capacitatea lor de a lua decizii centralizate. În anumite cazuri, înființarea lor poate fi situată într-un moment anume în timp, cum ar fi 4 iulie 1776 pentru Statele Unite ale Americii sau 24 octombrie 1945 pentru Organizația Națiunilor Unite. În alte cazuri, precum Franța sau Biserica Catolică, este mult mai potrivit să spunem că organizațiile au *evoluat* sau au *apărut*, și nu că au fost înființate. Cu toate că poate fi dificil să indicăm precis momentul în care o agregare slab structurată de indivizi se cristalizează într-un actor colectiv cu o identitate stabilă în timp, există totuși un moment în care ne dăm seama că a luat ființă.

Instituțiile sunt *moduri de a face lucruri* care apar sau evoluează gradual, cu toate că, odată ce au prins un anumit contur, ele pot fi adoptate conștient și implementate de către organizații. Să ne gândim, de pildă, la tranzițiile apărute între 1989-1990 în Europa de Est. Noile democrații au adoptat instituția economiei de piață și au creat burse de valori și alte organizații pentru implementarea instituției respective. În

constituțiile lor au pus la punct instituția Curții Constituționale, cu puterea de a respinge legile neconstituționale. În interacțiunea cu predecesorii lor autocrați, au adoptat instituții ale justiției de tranziție.

Aceste trei instituții, economia de piață, Curtea Constituțională și justiția de tranziție, nu au existat întotdeauna. Cu toate că Adam Smith nu greșea neapărat când enunța ca un fapt universal existența unei „anumite înclinații a naturii umane (...) de a face troc, de a schimba un lucru cu celălalt", această înclinație poate fi controlată de ceilalți. În special, poate fi blocată de neîncrederea care caracterizează „războiul fiecăruia împotriva tuturor", în starea naturală a lui Hobbes. Pentru a depăși neîncrederea, poate fi nevoie de instituții. În condiții primitive, trocul poate fi prea riscant dacă cei puternici sunt pur și simplu în măsură să își însușească prin forță proprietatea celor slabi fără să ofere nimic în schimb. Prin urmare, este mult mai bine să ții lucrurile pentru tine, să nu le vadă nimeni. Chiar când se depășește această problemă fundamentală prin supraveghere, lipsa de încredere și existența paraziților pot fi impedimente împotriva unei funcționări eficiente a piețelor. Dacă eu am o idee bună despre cum să produc un anumit produs mai ieftin, pot ezita să îmi dezvolt ideea dacă știu că alții o pot fura fără să suporte costul dezvoltării ei. Instituția sistemului de patente este un răspuns la această problemă. Dacă cei care se împrumută vor să își folosească terenul sau casa drept gaj, creditorii vor fi accesibili doar dacă pot sta liniștiți că proprietatea respectivă nu a fost pusă gaj de mai multe ori. Pentru a-i asigura, statul trebuie să creeze instituția cadastrului, pe care o pot consulta înainte să acorde împrumutul. Aceste instituții, care pot fi acum implementate într-o formă foarte elaborată de statele nou apărute, au ajuns la forma actuală timp de secole.

Astăzi poate părea foarte natural ca, în cazul în care o țară are o constituție, să aibă nevoie și de un mecanism politic care să respingă legile neconstituționale. În mod obișnuit, această sarcină („controlul judiciar") va cădea în seama unei Curți, fie că este o Curte mixtă, care are și alte atribuții, cum se întâmplă în Statele Unite, fie că este o Curte Constituțională specializată, cum se întâmplă în Europa continentală. Din punct de vedere istoric, însă, existau constituții înainte să apară controlul judiciar.[1] Deși Constituția Statelor Unite a creat o Curte Supremă cu autoritate asupra legislației statelor, aceasta nu era autorizată și să declare legile federale drept neconstituționale. În 1803, Curtea și-a arogat puterea în cazul Marbury *versus* Madison într-un mod unilateral. În 1854, Curtea Supremă din Norvegia și-a asumat prerogativa de a exercita controlul judiciar, care nu este menționat în Constituția țării. În 1971, Consiliul Constituțional din Franța s-a autorizat, împotriva intențiilor exprese ale au-

[1] În Olanda, încă nu există control judiciar. Controlul judiciar în Suedia este constrâns de prevederea constituțională că o lege trebuie să fie în mod „evident" neconstituțională pentru a putea fi respinsă de Curte. Probabil că astfel de legi nu ar fi adoptate de la bun început.

torilor constituției în vigoare, să respingă legislația care încălca Declarația Drepturilor Omului și Cetățeanului din 1789. Curtea Supremă a Israelului și-a acordat dreptul de a respinge legislația care încalcă una dintre Legile Fundamentale ale țării, transformând astfel constituția parțială a țării într-una completă. Cu toate că fiecare dintre aceste decizii a fost însoțită de controverse la momentul respectiv, toate, cu excepția celei din urmă, s-au consolidat în trăsături de necontestat ale peisajului instituțional.

Astăzi se așteaptă ca liderii regimurilor represive să fie trași la răspundere penală atunci când părăsesc poziția de conducere prin forță sau voluntar. Dacă noua putere nu ia singură această inițiativă, comunitatea internațională poate să o ia pentru ei (Ruanda și fosta Iugoslavie), poate exercita presiuni asupra regimului la putere (Indonezia) sau poate crea Curți hibride, care să cuprindă atât judecători din țara respectivă, cât și din alte state (Sierra Leone și Cambodgia). În ziua de azi există un fel de instituție numită justiție de tranziție. Însă, dacă ne uităm înapoi în istorie, execuțiile extra-judiciare, cu scopul răzbunării ori al incapacitării, au constituit maniera predilectă de a-i elimina pe liderii regimurilor opresive. Decizia aliaților de a-l exila pe Napoleon în 1814 și 1815 și de a nu-l ucide a fost un pas important în procesul de „încătușare a mâinii răzbunării". La sfârșitul celui de-al Doilea Război Mondial, mulți lideri aliați au vrut ca principalii criminali naziști să fie împușcați pe loc, ca niște criminali, fără a mai fi supuși unui proces. Într-un final, însă, procesul de la Nürnberg a manifestat un respect aproape exemplar pentru nevoia de a condamna pe baza unor probe și întrunind condițiile unui proces corect. Mai mult, unii dintre acuzați au fost chiar achitați. În unele țări ieșite de sub ocupația nazistă au existat elemente de justiție populară „sălbatică" între 1944 și 1945, dar acestea au fost rapid înlocuite de statul de drept. După căderea regimului comunist, 44 de ani mai târziu, a existat o înțelegere generală a faptului că practicile împotriva statului de drept ale fostelor regimuri nu trebuie imitate, nici chiar în judecarea fărădelegilor acestora.

Problema monitorizării

O organizație poate avea membri sau angajați. Membrii pot fi și angajați, cum este cazul cooperativelor muncitorești. Membrii interacționează orizontal prin procesele de argumentare, negociere și vot pe care le-am discutat în capitolul trecut. Relațiile verticale dintre angajați și superiori au un caracter diferit. Pentru a simplifica, să presupunem că organizațiile au un singur actor executiv („directorul") și mulți angajați („agenții"). O *problemă director-agenți* apare când, după cum se întâmplă adesea, directorul și agenții au interese diferite. Interesul muncitorilor poate fi acela de a avea un ritm de lucru moderat, în timp ce patronatul vrea ca ei să muncească mai mult. Dacă patronatul vrea să aducă interesele la un numitor comun plătindu-i pe muncitori mai mult pentru efortul lor, poate fi dificil sau costisitor să monitorizezi cât de mult muncesc aceștia. Tradițional, aceasta este sarcina șefului de echipă,

dar cum poate fi patronul sigur că șeful de echipă nu cere mită de la muncitori sau nu își folosește autoritatea pentru a-și promova propriile scopuri financiare sau sexuale? Locul de muncă caracteristic pentru secolul al XIX-lea era adesea definit ca „tirania șefului de echipă". În astfel de cazuri, cine îi supraveghează pe supraveghetori? După cum vom vedea în curând, aceste probleme pot apărea și când directorul și agenții au interese care coincid, cum se întâmplă în cooperativele muncitorești.

În mod asemănător, șeful unei agenții de stat are interesul ca angajații săi să fie cinstiți, adică să nu ceară și să nu ia mită de la cetățeni. Are, de asemenea, și un interes privitor la eficiență, astfel încât dimensiunea sectorului public să fie păstrată la un minim necesar. Interesele angajaților se pot manifesta în direcții opuse din ambele puncte de vedere. Dacă sunt motivați doar de propriul interes economic, aceștia vor lua mită ori de câte ori pot fără să fie prinși. Ca o consecință a interesului pentru putere, agenții sunt stimulați să crească dimensiunea departamentelor și numărul de subordonați. Din nou, monitorizarea se poate dovedi dificilă. Directorul poate uneori să surprindă un agent în timp ce ia mită, dar în general nu se poate baza pe această metodă. Directorul poate încerca să reducă oportunitățile în care se poate manifesta corupția organizând, de exemplu, licitații competitive pentru contractele publice, dar această precauție nu ajută dacă agenții concep contractele astfel încât acestea să-i favorizeze pe anumiți furnizori. Întrucât agenții au aproape un monopol asupra informației, directorii pot fi incapabili să își dea seama care cereri pentru suplimentări de personal sunt justificate și care nu.

Subordonații nu sunt singurii care beneficiază de stimulente care nu corespund fișei postului. Unii președinți ai universităților americane au fost obligați să demisioneze după ce a ieșit la iveală că își acordau salarii uriașe sau își remobilau casa cu banii organizației. Un vicepreședinte american (Spiro Agnew) a fost nevoit să demisioneze după ce s-a dovedit că era corupt. Cleptocrația – domnia hoților – a devenit un fenomen din ce în ce mai larg răspândit peste tot în lume. Cu toate că, într-un anumit sens, ei sunt directorii, astfel de lideri pot fi și ei supuși monitorizării. Însă supraveghetorii (votanții, consiliile de administrație, acționarii, Banca Mondială sau Fondul Monetar Internațional) au avut adesea insuccese notorii în reglementarea comportamentului directorilor generali sau șefilor de stat. Cum se întâmplă și în alte cazuri, le poate lipsi *informația* relevantă de care au nevoie pentru a corecta anumite excese sau *stimulentul* pentru a o face.

În cooperativele muncitorești poate apărea un conflict între muncitori ca agenți și muncitori ca directori. Adresându-se Congresului pentru Științe Sociale în 1863, Sir James Kay Shuttleworth spunea următoarele cu privire la morile de bumbac cooperatiste din Lancashire:

B.C.U. „M. EMINESCU" IAȘI

> S-a pus apoi următoarea problemă de mare importanță: Ce beneficii ar trebui să aibă acționarii din această moară, dincolo de profiturile obișnuite? Prima cerință înaintată în astfel de societăți a fost aceea că familiile acționarilor ar trebui să aibă prioritate în selectarea muncitorilor la moară. (...) Acesta fusese martor pe proprietatea sa la eșecul uneia dintre aceste preocupări. Exista o dorință de a introduce un principiu al cooperării, anume ca acționarii să aibă avantajul angajării propriilor familii în aceste mori. Efectul imediat a fost că, în loc să ducă o disciplină mai strictă și o atenție mai mare pentru funcționarea utilajelor, atât de necesare în morile de bumbac (și se poate menționa că disciplina unui regiment este inferioară ca strictețe în comparație cu aceea a unei mori de bumbac), la ședințele trimestriale și semestriale muncitorii făceau cele mai iritante plângeri împotriva supraveghetorilor, iar un supraveghetor care îndrăznea să dea afară un muncitor care era acționar se găsea în mare pericol de a fi el însuși dat afară la următoarea ședință.[1]

O altă problemă care apare destul de frecvent în cooperative se datorează reticenței de a renunța la membri în perioade în care cererea este scăzută. Comentând declinul întreprinderii Plate-Locksmiths din Woverhampton în 1878, un contemporan scria:

> Dacă afacerea ar fi fost condusă de un producător privat, probabil că acesta i-ar fi dat afară pe muncitorii care nu mai erau necesari odată cu scăderea cererii pentru încuietori și s-ar fi concentrat pe dezvoltarea comerțului rămas. Însă acest lucru presupunea, din partea corpului de muncitori care formau societatea, un sacrificiu de sine prea mare și pentru care nu erau pregătiți. Ei au optat să-și consolideze depozitele în speranța că va crește cererea. Însă aceasta nu și-a revenit înainte ca resursele să se epuizeze, iar ei au dat în cele din urmă de necaz. Datoriile s-au înmulțit și cei mai buni dintre muncitori au părăsit cooperativa.

Soluții la problema director-agent

Evitarea muncii, proliferarea angajărilor, corupția și alte probleme pot fi combătute fie modificând stimulentele agenților, fie oportunitățile acestora (capitolul 9). Ultima soluție este greu de pus în practică. Pentru a fi eficient, un agent are nevoie

---

[1] În 1875, directorul general de la Ouseborn Co-Operative Engine Works Company s-a plâns de lucruri foarte asemănătoare atunci când a dat o explicație pentru eșecul companiei: „Reglementările de orar de la Ouseborn, precum regula că orele suplimentare se adună în fiecare zi, restricțiile excesive asupra muncii din partea unor oameni, ponderea restrânsă a ucenicilor, discuțiile și delegațiile frecvente, cantitatea excesivă de muncă suplimentară necesară onorării contractelor și, în general, nevoia de disciplină rezultată în urma impresiei că fiecare este propriul său stăpân – toate au contribuit la creșterea costurilor de producție."

de o anumită independență și libertate de acțiune. Sclavii au fost doar rareori folosiți în meserii care presupun aptitudini și grijă. Se poate dovedi imposibil de a organiza lucrurile astfel încât agentul să poată urmări scopurile directorului obținându-se în același timp să și le urmărească pe ale sale. Am putea încerca să aproximăm acest ideal obligându-i pe decidenți să ocupe funcții pentru o perioadă suficient de mică de timp, astfel încât să fie greu ca aceștia să fie mituiți. Sistemul de juriați și colegiile electorale americane (în versiunea lor inițială) au fost justificate, printre altele, pe astfel de temeiuri. Numirea pentru o perioadă scurtă de timp a reprezentanților aleși (de multe ori, combinată cu neeligibilitatea lor pentru alt mandat) și rotația frecventă a reprezentanților numiți în funcție se presupune că ajută la reducerea oportunităților pentru corupție. Ultimele practici, însă, presupun costuri serioase pentru a fi eficiente. Funcționarii nici nu se obișnuiesc bine cu munca pe care trebuie să o desfășoare că sunt obligați să renunțe la post.

Pare mai promițător să găsim o soluție privitoare la stimulente. Recompensarea muncii asidue și penalizarea acelora care evită munca, de pildă, ar putea contribui la alinierea intereselor agenților cu acelea ale directorului. În meseriile în care efortul individual poate fi măsurat prin producție individuală, salarizarea în funcție de cantitatea produsă ar putea rezolva problema. Însă, uneori, tehnologia nu îngăduie unei firme decât să măsoare producția unui grup de muncitori. De asemenea, utilizarea salariilor în funcție de producție poate fi contraproductivă. Dacă îl plătești pe fiecare membru al unei echipe de fotbal (sau, cel puțin, pe atacanții ei) în funcție de numărul de goluri înscrise, întreaga echipă va funcționa mai puțin eficient. Un bun jucător de echipă este acela care îi pasează mingea cuiva aflat într-o poziție mai bună pentru a marca. Firma ar putea apoi să încerce o rezolvare a problemei stabilind un sistem de bonusuri pe echipă, astfel încât muncitorii pot avea un stimulent să se supravegheze unii pe alții.

Însă, cum se întâmplă adesea, în cazul în care costurile monitorizării sunt mult mai mari decât beneficiile aduse în plus persoanei care este responsabilă cu supravegherea, persoana respectivă nu are niciun interes *material* pentru monitorizare. Un jucător de fotbal ar putea considera că nu merită efortul care ar putea fi considerabil) să-l pună la locul lui pe coechipierul cu aere de vedetă. Cu toate că s-ar putea dezvolta o normă socială împotriva evitării muncii, datele empirice pe acest subiect sunt ambigue. Sunt șanse mult mai mari ca normele să se dezvolte la nivelul unor grupuri mici atât pentru că membrii lor se pot observa unii pe alții, cât și pentru că fiecare membru are mult mai mult de pierdut din partea comportamentului parazitar al celorlalți. Însă nici chiar aici aceste norme nu apar spontan (capitolul 22). Normele împotriva celor cu aere de vedetă sau împotriva celor care joacă egoist în meciurile de fotbal se pot datora strategiilor antrenorului, care are puterea să îi pedepsească pe jucătorii care au o părere prea bună despre sine și care desconsideră echipa. Se știe, însă, că unii jucători

își pedepsesc coechipierii care nu pasează refuzând la rândul lor să le paseze când ceilalți sunt într-o poziție mai bună pentru a marca.

Problema referitoare la cine îi păzește pe paznici sau cine îi supraveghează pe supraveghetori ar putea fi rezolvată teoretic prin ceva de felul unui *sistem circular*, în care fiecare persoană dintr-un lanț are un stimulent material direct pentru a pedepsi persoana din stânga sa întrucât, dacă nu o face, va fi pedepsită de persoana din dreapta sa. În practică, astfel de aranjamente nu par să existe. Jucătorii de fotbal pasează mingea celor care o pasează către cei care nu pasează. Tentația de a fi parazitar poate fi, însă, mijlocită de *încredere*. Dacă firmele oferă salarii peste oferta pieței, muncitorii (sau o parte a lor) pot răspunde depunând un efort suplimentar. Problema e că s-ar putea ca firma să nu aibă vreun stimulent pentru a iniția un astfel de sistem. Cu toate că producția totală (și, prin urmare, veniturile totale) va crește, și costurile salariale vor crește. Dacă rezultatul net îl constituie un nivel mai scăzut al profiturilor, firma se va afla într-o situație mai bună păstrând salariile competitive.

Organizarea neîncrederii

Constituțiile politice aspiră adesea să fie un sistem circular, în care să existe un echilibru instituțional în care performanța fiecărei instituții să fie monitorizată de alta. Această aspirație nu este realizată în toate constituțiile existente, poate chiar în niciuna. Mașinăria constituțională se bazează cel mai adesea pe un „motor nemișcat" sau un „verificator neverificat", o parte care monitorizează performanța tuturor celorlalte părți fără a fi ea însăși supusă monitorizării. În Statele Unite, această poziție aparține Curții Supreme. În alte țări, poate fi ocupată de parlament. În România, de exemplu, Parlamentul poate respinge deciziile Curții Constituționale dacă aceasta contestă legislația în vigoare (o moștenire din perioada comunistă, când Parlamentul era omnipotent din punct de vedere formal, dar *de facto* nu avea nicio putere). Și în Franța Parlamentul a fost motorul nemișcat până în 1971. Însă, cel puțin în tradiția anglo-americană, construcția constituțională se bazează pe principiul formulat de David Hume: „Este (...) o maximă politică dreaptă aceea că trebuie să presupunem că fiecare om este un ticălos." Din acest punct de vedere, constituțiile sunt o formă de neîncredere organizată.

Constituția americană este, poate, expresia cea mai elaborată a acestei atitudini. Printre cele mai frapante trăsături ale dezbaterilor din 1787 din cadrul Convenției Federale și ale documentului pe care aceasta l-a emis o reprezintă preocuparea permanentă față de potențialul pentru corupție, mită și amenințări și nevoia de a astupa toate fisurile prin care toate acestea ar putea să intre. Pentru a menționa doar un exemplu, Comitetul Special cu privire la reprezentare a recomandat, într-un raport din 9 iulie, ca, „din moment ce situația curentă a Statelor se va modifica atât din punctul de vedere al averii, cât și al numărului de locuitori, legislativul să fie autorizat

din când în când să mărească numărul de reprezentanți". Drept răspuns, Edmund Randolph se declara „temător în privința faptului că, din moment ce numărul avea să se modifice după bunul plac al legislaturii naționale, nu ar lipsi niciodată un pretext pentru a amâna modificările, astfel încât puterea să rămână în mâna acelora care o dețin". George Mason a argumentat în mod asemănător: „Putem fi siguri, având în vedere natura omului, că aceia care dețin puterea nu o vor ceda atâta timp cât o pot păstra." Am putea face, alături de Hume, afirmația mai slabă că trebuie *presupus* că aceia care dețin puterea vor încerca să o păstreze. Cu toate că unii politicieni pot lucra pentru binele comun (și toți vor pretinde acest lucru), trebuie să fim rezervați când vine vorba despre acordarea încrederii. Mai mult, ar trebui să presupunem că aceia care dețin puterea o vor folosi pentru a o reține *și spori*, dacă pot. Din acest punct de vedere, constituțiile sunt instrumente prin care aceștia sunt împiedicați să facă așa ceva. Nu este singurul lucru pe care îl fac constituțiile, dar este adesea unul dintre lucrurile pe care îl au în vedere.

Puterea acestui cadru normativ de a *explica* adoptarea constituțiilor reale depinde de cât de mult urmăresc aceia care construiesc acest cadru binele comun, și nu propriile interese și pasiuni. După cum am remarcat spre finalul capitolul 13, nu trebuie să luăm ca sigur faptul că aceștia nu pot fi afectați de interese și pasiuni. Însă presiunea normativă de a acționa *ca și cum* ar fi motivați de rațiune – forța civilizatoare a ipocriziei – poate fi mai puternică într-o adunare constituantă decât într-o legislatură ordinară. De asemenea, după cum am susținut în capitolul precedent, gradul de pătrundere a interesului personal în chestiunile care urmează a fi decise tinde să fie mai mic în cadrul constituant. Din acest motiv, adunările constituante care funcționează în spatele ușilor închise pot fi capabile să reducă efectul distorsionant al pasiunilor fără să plătească și costul unor negocieri excesive bazate pe interes.[1]

Dat fiind acest avertisment, în cele ce urmează voi schița un exemplu stilizat al modului în care neîncrederea organizată poate fi întruchipată într-o serie de prevederi legate între ele. Multe dintre aceste trăsături se găsesc în sistemele reale. Altele au fost incluse doar pentru a ilustra până unde trebuie să se meargă câteodată pentru a se închide toate posibilele fisuri.

---

[1] Nu sunt în întregime de acord cu această perspectivă negativă asupra pasiunilor. Hegel susține că „nimic măreț în lume nu a fost realizat fără pasiune". Tocqueville scrie că, în vremuri de criză, oamenii fie coboară sub, fie depășesc nivelul lor obișnuit. Mai mult, imparțialitatea nu exclude pasiunea, după cum au demonstrat-o membrii Adunării Constituante franceze în 1789, când erau „beți de dezinteres". Unul dintre cei mai impresionanți dintre aceștia, contele de Clermont-Tonnerrre, a supus că „anarhia este schimbare înspăimântătoare, dar necesară, și singurul moment în care se poate stabili o nouă ordine a lucrurilor. Măsurile uniforme nu se pot lua în momente de acalmie." Însă, cu toate că poate produce scopuri admirabile, entuziasmul poate să submineze gândirea limpede de care e nevoie pentru a le realiza.

1. Pentru a împiedica guvernul să nu se *implice în justiție politică*, constituția ar trebui fie să solicite jurii pentru procesele penale, fie să stipuleze ca judecătorii să fie numiți la întâmplare în procese (sau selectați de un soi de procedură mecanică). În absența unor clauze de acest tip, guvernul poate să intenteze proces împotriva adversarilor politici și să se asigure că aceștia vor fi condamnați prin numirea judecătorului care să o facă.
2. Pentru a împiedica guvernul să *manipuleze sistemul electoral*, legea electorală ar trebui prevăzută în constituție până în cele mai mici detalii și nu ar trebui lăsată pe seama determinării statutare. Reorganizarea districtelor electorale pentru a ține cont de schimbările în rândul populației se poate face fie printr-o formulă mecanică, fie poate fi încredințată unei comisii independente. Istoria americană și franceză recentă pune la dispoziție numeroase instanțe în care guvernele la putere au modificat regulile electorale pentru a-și menține puterea.
3. Pentru a împiedica guvernul să *manipuleze fluxul de informație*, radioul și televiziunea de stat ar trebui guvernate de comisii independente („modelul britanic"), nu de comisii numite de guvern („modelul francez"). Trusturile de presă private ar trebui protejate. Spațiul de emisie ar trebui alocat de comisii independente. Guvernului nu ar trebui să i se dea voie să raționalizeze resursele care sunt vitale pentru presă, precum hârtia și cerneala tipografică, sau astfel de decizii ar putea fi încredințate unui „comitet independent de raționalizare". Combinația dintre alegeri nemanipulate și presa independentă s-a dovedit o modalitate foarte eficientă de a împiedica foametea și alte dezastre. Dacă cei tentați să acționeze contrar interesului public știu că activitatea lor va fi dezvăluită electoratului, care poate apoi să îi dea jos, este mai puțin probabil că se vor angaja în astfel de comportamente.
4. Pentru a împiedica guvernul să *manipuleze politica monetară*, ca instrument pentru realegere, această politică trebuie încredințată unei bănci centrale independente. O alternativă mai puțin atractivă este aceea de a lăsa politica monetară să fie decisă de un program de calculator elaborat de trei economiști aleși la întâmplare din primele cinci departamente economice din țară, programul urmând a fi revizuit o dată la cinci ani.
5. Pentru a-i împiedica pe bancherii dogmatici și partizani ai băncii centrale să *elaboreze politici monetare dezastruoase*, ar trebui să fie posibilă demiterea lor în cazul în care o foarte mare majoritate (două treimi sau trei sferturi) ar decide astfel în parlament. Bancherii ar trebui să dețină mandate limitate în timp (spre deosebire de foștii guvernatori ai Băncii Centrale a

Italiei) și ar trebui să aibă un singur mandat (spre deosebire de guvernatorul italian din prezent).

6. Pentru a împiedica guvernul să *manipuleze informațiile statistice* ca instrument de realegere, culegerea de date ar trebui încredințată unui birou independent de statistică. În Marea Britanie, ministrul de finanțe a făcut pași importanți pentru a consolida încrederea în datele oficiale hotărând ca Biroul Național de Statistică să fie independent de guvern (*Financial Times*, 2 decembrie 2005). Conform unui titlu din *New York Times* (22 septembrie 2002), „unii experți se tem de influența politicului asupra agențiilor care se ocupă cu datele despre criminalitate". Au fost raportate, de asemenea, presiuni politice asupra statisticilor legate de piața muncii („Biroul de statistică a muncii ar trebui investigat", *New York Post*, 29 septembrie 2000). În cazuri mai subtile, alegerea unui model econometric de previziune poate fi și ea o mizã. Evident, cu cât oamenii cred că rata criminalității și a șomajului sunt mai scăzute și că perspectivele de creștere sunt mai bune, cu atât mai mari sunt șansele pentru politicienii în funcție să fie realeși. Pe lângă obstrucționarea guvernului de la a folosi politica monetară pentru a manipula realitatea, acesta ar trebui împiedicat să folosească statistica pentru a manipula credințele despre realitate.
7. Pentru a împiedica guvernul să *înfometeze opoziția*, partidele politice majore ar trebui să primească subvenții în funcție de proporția votanților și în funcție de proporția numărului de parlamentari. Nivelul subvențiilor ar trebui fixat prin constituție și ar putea fi exprimat prin prețul aurului.[1]
8. Pentru a împiedica guvernul să *elaboreze legislație în interes propriu*, constituția ar putea să stipuleze că anumite categorii de legi pot intra în vigoare doar după o anumită perioadă de timp (un an sau mai mult) după ce sunt adoptate. Prin acest mecanism s-ar crea un văl de ignoranță artificial, în spatele căruia chiar și partidele motivate doar de interesul propriu ar fi determinate să ia măsuri imparțiale.
9. Pentru a împiedica guvernul *să ocolească aceste restricții* folosindu-se de majoritatea parlamentară ca să aducă amendamente constituției, amendamentele ar trebui să aibă nevoie de majoritate aproape absolută, de amânări sau de ambele strategii. Clauza care specifică aceste lucruri cu privire la amendamente ar trebui să fie imună la orice formă de amendament.

---

[1] O dorință asemănătoare de a face constituția pe cât posibil explicită și nemanipulativă a stat și în spatele ideii lui Mason de a impune salarii fixe pentru judecătorii federali „luând ca standard grâul sau un alt bun a cărui valoare este permanentă".

10. Pentru a împiedica guvernul să *ignore aceste restricții*, după cum s-a întâmplat până la un moment dat în Franța, constituția ar trebui să stipuleze controlul judiciar exercitat de o Curte Constituțională independentă sau de Curtea Supremă.
11. Pentru a împiedica guvernul să manipuleze controlul judiciar prin creșterea numărului de judecători de la Curtea Constituțională sau Curtea Supremă, acel număr nu ar trebui lăsat pe seama statutului, după cum este cazul în Statele Unite și alte țări, ci ar trebui specificat în constituție.
12. Pentru a împiedica judecătorii dogmatici și partizani ai Curții Constituționale sau ai Curții Supreme să *ignore mari majorități populare*, ar trebui să fie posibilă demiterea lor dacă există o mare majoritate (două treimi sau trei sferturi) în favoarea unei astfel de măsuri în parlament. Spre deosebire de membrii Curții Supreme americane, mandatul lor ar trebui să fie limitat. De asemenea, aceștia ar trebui să fie neeligibili pentru un nou mandat.

După cum indică acest exemplu, o constituție poate fi o mașinărie foarte complicată. Una dintre piese poate lipsi, altele pot ajunge ineficiente sau dăunătoare (vezi discuția de mai încolo). În același timp, pentru constituție există un adevărat pericol ca aceasta să fie atât de rigidă încât să împiedice sistemul politic să răspundă la urgențe. Constrângerile constituționale asupra politicii monetare sau asupra legii electorale ar putea, în anumite cazuri, să producă dezastre. La Convenția Federală din Philadelphia, George Mason remarca: „în ciuda faptului că uram de moarte banii de hârtie, totuși nu puteam prevedea toate urgențele care ar putea apărea, deci nu eram dispus să leg mâinile legislativului. Am observat că ultimul război nu s-ar fi putut purta dacă ar fi existat o astfel de interdicție." Asemănător, în 1946, Parlamentul italian a decis împotriva reglementării stabilității monetare prin Constituție, aducând ca obiecție nevoia guvernului de a fi liber să acționeze în vreme de război. În 1945, de Gaulle a manipulat sistemul electoral francez pentru a nu le da voie comuniștilor să ajungă la putere. Comentând mulți ani mai târziu asupra unei propuneri de a include legea electorală în Constituție, acesta a spus că „nu se poate ști niciodată ce urmează să se întâmple. Ar putea exista la un moment dat motive pentru a reveni la votul proporțional de dragul interesului național, cum s-a întâmplat în 1945. Nu ar trebui să fim legați de mâini." Privind retrospectiv, Franța ar trebui să fie bucuroasă că legea electorală nu a fost inclusă în Constituție.

Problema poate fi rezumată în două observații faimoase cu privire la Constituția americană. Comentând Legea cu privire la Ku Klux Klan din 1871, John Potter Stockton a spus că, de fapt, „constituțiile sunt lanțuri cu care oamenii se leagă în momentele lor de luciditate pentru a nu muri de o mână sinucigașă în momentele de

nebunie". În 1949, Robert Jackson susținea că „Declarația Drepturilor Omului nu este un pact suicidal". Nu există nicio metodă simplă de a ne asigura că o constituție menită ca instrument de împiedicare a sinuciderii nu devine ea însăși un pact suicidal.

Problema celei de a doua opțiuni

Trecând peste această chestiune, să presupunem acum că, dacă cele douăsprezece condiții enumerate (sau altele cu caracter general) sunt satisfăcute, sistemul politic va deveni *optim* într-un anumit sens, pe care trebuie încă să îl descriem. În lumea reală, acest rezultat este puțin probabil că va fi atins. Este greu să faci totul ca la carte. Însă am putea dori să aproximăm acest optim cât mai mult posibil, pornind de la presupoziția aparent rezonabilă că, de fapt, cu cât sunt satisfăcute mai multe condiții pentru optimalitate, cu atât ne apropiem mai mult de acel optim. Această presupoziție este falsă. În condiții foarte generale, nu este adevărat că o situație în care multe, dar nu toate, dintre condițiile ca un optim să fie atins sunt satisfăcute devine în mod necesar sau are șanse mari să devină superioară unei situații în care sunt satisfăcute un număr mai mic de astfel de condiții. Revenind la cele douăsprezece instrumente citate mai devreme, nu este clar a priori dacă un controlul judiciar fără verificarea judecătorilor este mai bun decât un sistem fără control judiciar sau că băncile centrale independente fără control asupra bancherilor sunt de preferat unui sistem care lasă politica monetară la latitudinea guvernului.

Discuția lui Tocqueville despre Vechiul Regim (*Ancien Régime*) din Franța poate fi citită din această perspectivă. Acel sistem era caracterizat de un număr de trăsături care ar fi absente într-o societate bine așezată. Administrația regală avea puteri vaste, arbitrare și nedefinite. Corupția funcționarilor făcea imposibilă o birocrație rațională. Obstrucția produsă de acele *parlements*, Curți puternic politizate, care acționau mai mult pentru servirea propriilor interese, făcea dificilă implementarea unor politici consistente. Însă, argumentează Tocqueville, dată fiind prima trăsătură, prezența celorlalte două devenea de fapt benefică:

> Guvernul, în dorința sa de a transforma totul în bani, a scos la vânzare mai întâi funcțiile publice, privându-se pe sine de capacitatea de a le acorda și revoca după propria voință. Una dintre pasiunile sale a afectat astfel puternic succesul unei alte pasiuni: lăcomia sa a funcționat în sens contrar ambiției. Pentru a putea acționa, s-a văzut în permanență redusă la folosirea unor instrumente pe care nu le-a creat ea însăși și pe care nu le putea distruge. Prin urmare, dorințele sale cele mai puternice primeau o îndeplinire foarte slabă. Constituția bizară și plină de defecte a funcțiilor publice a luat locul oricărei garanții politice împotriva omnipotenței guvernului central. Era un tip de barieră ciudată și prost construită, care bloca puterea guvernului și îi atenua impactul. (...) Intervenția neregulată a Curților în actul de guvernare, care întrerupea adeseori buna administrare a

treburilor, servea în anumite momente pentru a proteja libertatea: era un mare rău care punea limite unuia și mai mare.

În analiza pe care o face legislației engleze din secolul al XVIII-lea, James Fitzjames Stephen remarcă faptul că excesiva scoatere în evidență a detaliilor tehnice ale legii „a atenuat severitatea excesivă a vechii legi penale, însă într-un mod irațional și capricios". Comentând dreptul penal islamic, acesta scria că „a atenuat duritatea extravagantă a prevederilor sale prin reguli cu privire la probe care, practic, exclud posibilitatea aplicării respectivelor prevederi". Asemănător, Imperiul Otoman, regimul țarist în Rusia, Italia lui Mussolini și Spania lui Franco au fost caracterizate ca „regimuri despotice temperate de incompetență". În comparație cu eficiența implacabilă a Germaniei naziste, aceste regimuri au fost, de fapt, benigne.

Tocqueville observă, de asemenea, că, deși funcționarii Vechiului Regim au încercat să înlocuiască detestata *corvée* (munca forțată la drumuri) cu o taxă pe care să o folosească pentru întreținerea drumurilor, ei au renunțat la idee de teamă că, „odată stabilit acest fond, nimic nu mai împiedica Trezoreria să direcționeze banii spre propriile sale scopuri, rezultatul fiind acela că, în curând, cetățenii urmau să se aleagă atât cu o nouă taxă, cât și cu datoria muncii forțate". Acest fenomen este larg răspândit. Când politicienii propun ca spitalele finanțate din bani publici să dea prioritate la coadă indivizilor care pot fi ajutați pentru a reveni în rândul forței de muncă, aceștia promit de obicei să canalizeze câștigurile sociale astfel obținute spre spitale, astfel încât alți pacienți să beneficieze de acestea. Administratorii spitalelor tind să fie sceptici în fața unor astfel de propuneri, suspectând că doar bugetul guvernului poate beneficia de pe urma respectivelor câștiguri. În California, propunerea de a distribui apa fermelor care o pot folosi cel mai eficient pentru a genera un surplus, care urmează a fi folosit pentru a îmbunătăți sursa de apă și pentru ceilalți fermieri, a întâmpinat un scepticism asemănător. O situație cu două elemente suboptime – un sistem ineficient de acordare a priorității pentru alocarea bunurilor rare și un guvern care nu își respectă promisiunile (sau pe acelea ale predecesorilor) – poate fi mai bun decât unul în care primul element este îndepărtat.

Toate aceste cazuri seamănă cu ce se întâmplă când politicienii decid eliminarea animalelor care sunt o pacoste pentru populația umană pentru a descoperi ulterior că o pacoste și mai mare este reprezentată de vietățile pe care primele le țineau sub control. Astfel, după ce Mao Zedong a decis decimarea vrăbiilor pentru că mâncau grânele, a fost obligat să le reimporte din Uniunea Sovietică deoarece dăunătorii pe care vrăbiile îi țineau sub control au proliferat, cu rezultate ecologice catastrofale. Societățile, la fel ca sistemele ecologice, pot avea trăsături aparent absurde sau nocive a căror îndepărtare poate crea un rău și mai mare. Probabil din acest motiv, printre

altele, Edmund Burke și urmașii săi au fost atât de hotărâți în critica pe care au adus-o sistemelor instituționale raționaliste.

Acest mod de gândire poate fi dus prea departe. Odată ce muncitorii și-au câștigat dreptul de vot, oponenții au argumentat că eliminarea unei nedreptăți împotriva muncitorilor poate destabiliza societatea în moduri dintre cele mai diverse și poate crea nedreptăți încă și mai mari față de bogați, întrucât se aștepta ca averea lor să fie confiscată de săraci. Acest argument trasează semnul de egalitate între două seturi de considerații care nu au aceeași greutate. Dar, întrucât aceasta este o carte despre explicația fenomenelor sociale, nu despre evaluarea lor din punct de vedere normativ, nu voi continua discuția pe această temă.

◈◈◈

Notă bibliografică

Preiau distincția dintre organizații și instituții din C. Offe, „Institutions' role in the distribution and control of social power", în I. Shapiro, S. Skowronek și D. Gavlin (editori), *The Art of the State* (New York University Press, 2006). „Ascensiunea pieței" este o temă centrală în D. North, *Structure and Change in Economic History* (New York, Norton, 1981). Apariția controlului jurisdicțional este tema lui E. Smith (ed.), *Constitutional Justice under Old Constitutions* (Haga, Kluwer, 1995). Pentru apariția justiției tranziționale, vezi G.J. Bass, *Stay the Hand of the Vengeance* (Cambridge, MA, Harvard University Press, 2001). O lectură utilă asupra organizațiilor (așa cum le numesc uneori în mod idiosincratic) este C. Ménard (ed.), *The Political Economy of Institutions* (Cheltenham, Edward Elgar, 2004). Problema alinierii stimulentelor individuale și organizaționale este subiectul cărții lui J.-J. Laffont și J. Tirole, *A Theory of Incentives in Procurement and Regulation* (Cambridge, MA, MIT Press, 1994). Un manual cuprinzător despre corupție este A. Heidenheimer, M. Johnston și V. LeVine (editori), *Political Corruption* (New Brunswick, NJ, Transaction Publishers, 1989). Referințele la cooperativele engleze din secolul al XIX-lea provin din B. Jones, *Co-operative Production* (Oxford University Press, 1894; New York, Kelley, 1968). Ideea unui „sistem circular" de supraveghere mutuală este discutată în E. Kandel și E. Lazear, „Peer pressure and partnership", *Journal of Political Economy*, 100 (1992), pp. 801-807, și în R. Calvert, „Rational actors, equilibrium, and social institutions", în J. Knight și I. Sened (editori), *Explaining Social Institutions* (Ann Arbor, University of Michigan Press, 1995) (retipărit în Ménard, *The Political Economy of Institutions*). Importanța relativă a încrederii și stimulentelor în firme este discutată în E. Fehr și A. Falk, „Psychological foundation of incentives", *European Economic Review*, 46 (2002), pp. 687-724. Originile humeene ale Constituției americane sunt trasate în

M. White, *Philosophy*, The Federalist, *and the Constitution* (Oxford University Press, 1987). Citatul din de Gaulle este luat din A. Peyrefitte, *C'était de Gaulle*, vol. I (Paris, Fayard, 1994), p. 452. Ideea celei de a doua opțiuni a fost introdusă în R.G. Lipsey și K. Lancaster, „The general theory of the second best", *Review of Economic Studies*, 24 (1956), pp. 11-32. Discut analiza adusă de Tocqueville Vechiului Regim în „Tocqueville on 1789: Preconditions, precipitants, and triggers", în C. Welch (ed.), *The Cambridge Companion to Tocqueville* (Cambrige University Press, 2006). Pasajele citate din Stephen provin din *History of the Criminal Law of England* (Londra, Macmillan, 1883; Buffalo, NY, Hein, 1964), vol. I, p. 284, și vol. 3, p. 293. Referința la Mao Zedong se află în J. Chang și J. Halliday, *Mao: the Unknown Story* (New York, Knopf, 2005), pp. 430-431. Îmi prezint punctul de vedere despre bazele normative ale schimbării instituționale în capitolul 4 al cărții *Solomonic Judgments* (Cambridge University Press, 1989).

BCU IASI/CENTRAL UNIVERSITY LIBRARY

CONCLUZIE: ESTE POSIBILĂ ȘTIINȚA SOCIALĂ?

Ce putem considera drept știință?

Pentru a răspunde întrebării dacă o știință socială este posibilă sau dacă există deja trebuie să ne uităm la criteriile care trebuie îndeplinite pentru a numi o anumită activitate știință. Un criteriu extern ar putea fi următorul: o disciplină a devenit știință dacă: (1) aceia care o practică au atins consensul cu privire la ce este adevărat, fals, conjectural și ce este necunoscut în domeniul respectiv; (2) există un proces de progres cumulativ prin care teoriile și explicațiile, odată abandonate, sunt abandonate pentru totdeauna; (3) conceptele principale și teoriile pot fi exprimate în termeni clari și suficient de expliciți pentru a fi înțeleși de oricine este dispus să dedice timpul și efortul necesare acestei activități; și „clasicii" disciplinei sunt citiți doar de istoricii științei. După cum spunea Alfred Whitehead, „o știință care ezită să își uite fondatorii este pierdută".

Această descriere are intenția de a surprinde în mare statutul științelor naturii contemporane. Criteriul (1) nu este propriu-zis și pe deplin satisfăcut de niciuna dintre aceste discipline. Există controverse, de pildă, cu privire la teoria stringurilor, la echilibrul punctat sau la măsura în care creșterea sau natura sunt importante în dezvoltarea umană. Însă amploarea neînțelegerilor este exagerată de cei din afara comunității științifice relevante, de persoane care nu pot aprecia cantitatea imensă de cunoaștere asupra căreia există consens și care formează fundalul tuturor controverselor, dar și de persoane care pot confunda uneori caracterul provizoriu al unor cercetări cu caracterul categoric al unor teze. Criteriul (2) este în mod obișnuit satisfăcut. Există foarte puține inversări ale istoriei în științele naturii, nu există revolte ale neonewtonienilor împotriva lui Einstein sau revolte ale neolamarckienilor împotriva lui Darwin. Transformarea teoriilor semiobscure sau ciudate în primele lor formulări în material ușor asimilabil în manuale certifică satisfacerea criteriului (3). Cu rare excepții, de obicei în biologie (Darwin, d'Arcy Wentworth Thompson, Claude Bernard), criteriul (4) este și el satisfăcut.

BCU CLUJ / UNIVERSITY LIBRARY

Folosind aceste criterii, voi încerca în cele ce urmează să propun o evaluare a științelor sociale slabe, calitative și cantitative. După câteva observații în treacăt asupra științelor sociale slabe, voi argumenta în favoarea științei sociale calitative. Voi argumenta, într-un mod ceva mai discutabil, împotriva științei sociale cantitative, cel puțin așa cum este întâlnită aceasta în formele ei mai remarcabile. Susținătorul științei sociale calitative este, prin urmare, pus în situația incomodă de a purta un război pe două fronturi, înfruntând riscul constant de a fi acuzat de un oponent că este în cârdășie cu celălalt.

Științele sociale slabe

Științele sociale „slabe" au mai multe elemente în comun cu anumite forme de critică literară (sau cu literatura) decât cu investigațiile empirice calitative. Postmodernismul, teoria postcolonială, teoriile subalterne, deconstructivismul, psihanaliza kleiniană sau lacaniană și alte teorii au fost demascate pentru caracterul lor obscur de mulți autori, probabil cel mai eficient fiind Alan Sokal. După cum observă acesta într-un interviu, lipsa unui limbaj comun pentru o discuție rațională face imposibilă criticarea directă a acestor pseudoteoreticieni. Ai, însă, la îndemână posibilitatea să-i prinzi cu mâța-n sac, cum a procedat Sokal într-un mod foarte eficient prin trimiterea spre publicare a unui articol despre hermeneutica gravitației cuantice plin până la refuz de jargon lipsit de orice sens, dar care suna suficient de pompos.

Cu toate că membrii acestor cercuri pot dezvolta un stil de discuție care poate sugera un anumit grad de satisfacere a criteriului (1), acesta se reduce la o formă de pseudointersubiectivitate, în care reflexele verbale comune trec drept consens rațional. Criteriile (2) – (4) sunt doar arareori satisfăcute. Nu pare că există idei care să fi fost abandonate definitiv. Dar poate cel mai mult frapează nesatisfacerea criteriului (3). Derrida poate uimi prin limbajul său, dar „învățăturile" sale nu pot fi *predate*. Nu există manuale deconstructiviste despre „cum se face" (cu toate că există destule treceri în revistă sau „cartografieri" și „negocieri"), în bună măsură pentru că practicanții teoriei au tendința de a avansa în argumentare prin insinuări și întrebări retorice, în loc să riște aserțiuni categorice. Poate că setul de parodii al lui Frederick Crew, *Postmodern Pooh*, se apropie cel mai mult de idealul de manual. Cât privește criteriul (4), cultul înaintașilor, cu referințe obligatorii și exegeze, pare a fi obligatoriu. (În acest paragraf mă adresez, în mod evident, acelora deja convertiți.)

Știința socială calitativă

În cadrul științei sociale calitative includ cea mai mare parte a scrierilor istorice, precum și lucrări care iau forma „studiilor de caz", și nu a studiilor generale. Cred că cel mai bun mod de instruire pentru un cercetător în științele sociale este să

citească mult și profund istoria, alegând opere pentru calitatea argumentului, și nu pentru importanța sau relevanța subiectului. Câteva modele ar fi: James Fitzgerald Stephen, *A History of the Criminal law of England*, E.P. Thompson, *The Making of the English Working Class*, G.E.M. de Ste Croix, *The Class Struggles in the Ancient Greek World*, Joseph Levenson, *Confucian China and Its Modern Fate*, Paul Veyne, *Le pain et le cirque*, G. Lefebvre, *La grande peur*, Keith Thomas, *Religion and the Decline of Magic*, Tocqueville, *L'ancien régime et la Révolution*, Max Weber, *Agrarverhältnisse im Altertum*, Gordon Wood, *The Radicalism of the American Revolution*, Jean Egret, *La pré-révolution française*, Denis Crouzet, *Les guerriers de Dieu*, și Martin Ostwald, *From Popular Sovereignty to the Sovereignty of Law*. (Îmi asum anumite riscuri întrucât unele titluri din listă nu sunt considerate încă lucrări clasice.) Ceea ce acești autori și alții de același calibru au în comun este o combinație între autoritatea cu privire la chestiunile de fapt și atenția pentru generalizările potențiale, precum și posibilele contraexemple pentru generalizări. Dimensiunea cunoașterii lor le permite să extragă „detaliul relevant", precum și „anomalia robustă", furnizând astfel un control fundamentat atât în stimuli, cât și în realitate împotriva unor generalizări hazardate.

Același lucru este valabil despre autorii de „studii de caz", unul dintre cele mai importante fiind și în continuare *Despre democrație în America* a lui Tocqueville. Cu toate că depășește limitele acestei categorii, aș include și cartea lui Joseph Schumpeter, *Capitalism, Socialism, and Democracy* (*Capitalism, socialism și democrație*). O altă carte, din câte se pare excentrică, dar, cred, foarte convingătoare, este *Travels in France* a lui Arthur Young, în care acesta acoperă anii 1787-1789. Acestea sunt „portrete de caracter" ale unor întregi societăți sau regimuri și toate oferă o perspectivă comparativă. Cartea lui Marc Bloch, *La société féodale* (*Societatea feudală*), se numără și ea în rândul acestora. Cartea lui Alexandr Zinoviev, *The Yawning Heights* (*Culmile abisale*), nu este chiar un portret de caracter al comunismului sovietic, ci o caricatură în sensul bun al termenului: elimină aspectele neesențiale și scoate în evidență trăsăturile fundamentale exagerându-le. O versiune bine documentată este reprezentată de S. Coutrois *et al.*, *La livre noir du communisme* (*Cartea neagră a comunismului*). Trilogia lui Richard Evans despre cel de-al Treilea Reich reprezintă, îndeosebi pentru regimul nazist, ce înseamnă cartea lui Robert Paxton *What Is Fascism?* (*Ce este fascismul?*) pentru regimul fascist în general. Cartea lui Richard Bosworth, *Mussolini and Mussolini's Italy* (*Mussolini și Italia lui Mussolini*), citită împreună cu cărțile lui Evans, oferă intuiții frapante cu privire la diferența dintre un regim al cărui rău, cu toate că real, era de o anvergură mai redusă în comparație cu un regim care era rău până la măduvă.

Analiza istorică și studiile de caz nu furnizează doar materia primă pentru generalizări, ci conțin adesea generalizări implicite, pe care autorul le lasă la latitudinea cititorului. De pildă, Tocqueville este un maestru al generalizărilor latente. Din cauza unei combinații de aroganță a istoricului și aroganță a aristocratului, acesta nu a

binevoit să expliciteze eșafodajul teoretic, ci s-a prefăcut că merge cu pași mărunți. Cu toate că *Despre democrație în America* se bazează adesea pe schema dorință-oportunitate pentru a explica comportamentul, această schemă nu este explicitată în atât de multe cuvinte. Cu toate că *L'ancien régime et la Révolution* (*Vechiul Regim și Revoluția*) se bazează mult pe distincția dintre invidie și ură în analiza relațiilor de clasă, cititorii trebuie să își dea seama singuri care sunt cauzele și efectele sistematice ale acestor două emoții.

În categoria de știință socială calitativă intră și scrierile care fac generalizări *explicite*. În discutarea acestora nu voi face apel la citarea de autori sau opere, ci mă voi întoarce la paradoxurile pe care le-am enumerat în Introducere. Foarte important, toate „răspunsurile" se supun celei de-a doua avertizări pe care am menționat-o în Introducere: satisfac doar condiția minimă de a-și implica logic explanandumul. În anumite cazuri, aș putea risca să spun că explicația pe care o ofer este, probabil, cea corectă, dar în altele nu.

I ⇝ Gândirea

- De ce cred unii jucători la ruletă că, după o serie de cinci apariții consecutive ale roșului, este mai probabilă apariția roșului decât a negrului? *Răspuns*: pentru că sunt influențați de euristica disponibilității.

- De ce cred alți jucători la ruletă că, după o serie de cinci apariții consecutive ale roșului, este mai probabilă apariția negrului decât a roșului? *Răspuns*: pentru că sunt influențați de euristica reprezentativității.

- De ce, uneori, ni se schimbă preferințele prin simpla trecere a timpului? *Răspuns*: fie din cauza deprecierii hiperbolice, fie din cauza timpului de înjumătățire scurt al emoțiilor care le-au generat.

- De ce mulți dintre aceia care par a crede în viața de apoi vor ca aceasta să survină cât mai târziu posibil? *Răspuns*: deoarece credința lor nu este genul de credință care servește ca premisă pentru acțiune, ci doar le produce o anumită pace sufletească ori de câte ori se gândesc la moarte.

- De ce sunt oamenii reticenți, față de sine și față de ceilalți, în a recunoaște că sunt invidioși? *Răspuns*: pentru că sunt preocupați de imaginea de sine și pentru că, în anumite societăți, invidia este situată pe ultimele locuri în ierarhia normativă a motivațiilor.

- De ce sunt oamenii reticenți, față de sine și față de ceilalți, în a recunoaște că nu cunosc un anumit lucru? *Răspuns*: pentru că ființele umane sunt

animale care caută ordinea în lucruri și pentru că faptul de a mărturisi propria neștiință asupra chestiunilor importante produce disconfort psihic.

- De ce, în secolul al XVI-lea, credința convertiților la calvinism că oamenii sunt predestinați să ajungă sau în Iad, sau în Rai aducea mai multă împăcare sufletească decât credința că mântuirea poate fi atinsă prin fapte bune? *Răspuns*: pentru că gândirea magică i-a făcut să creadă că, prin aderiunea la Biserică, îl pot determina pe Dumnezeu să-i fi ales deja printre cei mântuiți.

- De ce este (uneori) adevărat că „cel care face rău nu poate să ierte"? *Răspuns*: pentru că indivizii aroganți sunt atât de reticenți în a admite că au făcut un rău, încât inventează motive pentru care persoana ofensată merita și continuă să merite răul făcut.

- De ce este rușinea mai importantă decât vina în anumite culturi? *Răspuns*: pentru că o societate care nu a conceptualizat vina va manifesta mai rar comportamente legate de vină.

- De ce victoria Franței la Campionatul Mondial de Fotbal din 1998 a provocat atât de multă bucurie în țară și de ce faptul că echipa nu s-a calificat din grupe în 2002 a provocat atât de multă mâhnire? *Răspuns*: pentru că elementul surprizei amplifică atât emoțiile pozitive, cât și pe acelea negative.

- De ce unele femei trăiesc un sentiment de rușine în urma unui viol? *Răspuns*: pentru că victimele suferă de o atitudine de „învinuire a victimei", care decurge din credința că lumea este esențialmente dreaptă.

- De ce ritualurile de inițiere umilitoare sporesc loialitatea față de grupul în care are loc inițierea? *Răspuns*: deoarece credința că ai îndurat o durere foarte mare ar crea degeaba o disonanță cognitivă.

II ↝ Acțiunea

- De ce, în zilele noastre, publicul aplaudă în picioare spectacolele de pe Broadway mai frecvent decât acum douăzeci de ani? *Răspuns*: pentru că spectatorii trebuie să simtă că primesc ceva de valoare în schimbul banilor cheltuiți.

- De ce pedepsele, în loc să reducă rata de incidență a comportamentului pe care îl vizează, mai degrabă o sporesc? *Răspuns*: pentru că utilizarea unor pedepse dure determină ură, care poate contrabalansa frica.

- De ce sunt oamenii refractari la încălcarea regulilor autoimpuse, chiar când respectarea lor nu are sens? *Răspuns*: pentru că le este teamă că o singură excepție ar putea distruge regula.

- De ce răzbunarea urmează mai degrabă modelul „Doi ochi pentru un ochi" decât „Un ochi pentru un ochi"? *Răspuns*: din cauza aversiunii față de pierdere.

- De ce randamentul pe termen lung al acțiunilor este mai mare decât al obligațiunilor (adică de ce valoarea acțiunilor nu crește astfel încât să egaleze randamentele)? *Răspuns*: din cauza unei combinații între aversiunea față de pierdere și „miopia deciziei".

- De ce rata sinuciderilor scade atunci când medicamentele periculoase se vând în folii, față cu vânzarea lor în flacoane? *Răspuns*: pentru că multe dorințe sunt atât de efemere încât desfacerea foliei ia suficient de mult timp pentru a-i face pe cei cu tendințe sinucigașe să se răzgândească.

- De ce niciunul dintre cei treizeci și opt de martori care au asistat la uciderea în bătaie a lui Kitty Genovese nu a chemat poliția? *Răspuns*: pentru că fiecare dintre ei a crezut că, din moment ce nimeni altcineva nu intervine, situația nu este atât de gravă.

- De ce unii oameni au adăpostit sau salvat evrei în timpul regimului nazist? *Răspuns*: pentru că li s-a cerut asta și le-a fost rușine să spună nu.

- De ce a vrut președintele Chirac alegeri anticipate în 1997 doar pentru a pierde majoritatea parlamentară? *Răspuns*: pentru că nu a prevăzut faptul că votanții vor trage concluzii din anunțul alegerilor anticipate care îi vor face să voteze împotriva lui.

- De ce unii părinți divorțați sunt dispuși să împartă custodia copilului în ciuda faptului că soluția preferată ar fi custodia totală, pe care ar putea chiar să o obțină dacă ar merge la tribunal? *Răspuns*: din cauza aversiunii față de risc.

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

- De ce este mai puțin probabil ca oamenii săraci să emigreze? *Răspuns*: pentru că nu își permit costurile cu transportul și nu se pot folosi pe sine ca gaj pentru un împrumut.

- De ce unii oameni economisesc în conturi speciale pentru Crăciun, conturi care nu oferă nici dobândă, nici posibilitatea efectuării unei retrageri înainte de Crăciun? *Răspuns*: pentru că știu că, dacă economisesc în modul obișnuit, deprecierea hiperbolică i-ar putea determina să scoată banii mai devreme.

- De ce se angajează oamenii în proiecte cu valoare scontată negativă, cum ar fi construcția aeronavei *Concorde*? *Răspuns*: din cauza orgoliului sau din cauza aversiunii față de risc.

- De ce în cadrul „justiției de tranziție" (care presupune că agenții responsabili pentru un regim autocrat sunt aduși în fața unui tribunal odată cu trecerea la democrație) cei judecați mai devreme primesc pedepse mai severe în comparație cu cei judecați mai târziu? *Răspuns*: din cauza timpului de înjumătățire foarte scurt al emoției răzbunării.

- De ce, în piesa omonimă a lui Shakespeare, Hamlet își amână răzbunarea până în ultimul act? *Răspuns*: pentru că Hamlet are voința slabă și tensiunea nu putea fi rezolvată înainte de finalul piesei.

III ⇝ Lecții desprinse din științele naturii

- De ce este mai probabil ca părinții să își ucidă copiii adoptați sau vitregi, și nu copiii biologici? *Răspuns*: pentru că doar cei din urmă le pot transmite genele mai departe.

- De ce este atât de rar întâlnit incestul între frați, în condițiile în care tentația există, iar oportunitatea de asemenea? *Răspuns*: pentru că selecția naturală a favorizat un mecanism care inhibă dorința sexuală pentru membrii familiei de aceeași vârstă și de sex opus.

- De ce sunt oamenii dispuși să investească în proiectele altora în condițiile în care cei din urmă păstrează întregul profit? *Răspuns*: pentru că selecția de grup a favorizat tendința de cooperare.

- De ce uneori oamenii se răzbună suportând costuri materiale, dar fără niciun beneficiu material? *Răspuns*: pentru că selecția de grup a favorizat tendința de pedepsire a celor care nu cooperează.

BCU CLUJ-NAPOCA UNIVERSITY LIBRARY

- De ce oamenii trag concluzii care nu sunt susținute de dovezi? *Răspuns*: pentru că selecția naturală a încurajat comportamentul de a căuta un model în lucruri.

IV ↝ Interacțiunea

- De ce susținătorii unui partid socialist votează uneori cu comuniștii, împiedicând astfel propriul partid să câștige? *Răspuns*: pentru că fiecare dintre ei este victima „sindromului fratelui mai mic", care îi împiedică să își dea seama că și alții ar putea face același lucru.

- De ce unele state care de abia și-au câștigat independența adoptă ca limbă oficială limba foștilor asupritori imperialiști? *Răspuns*: pentru că într-o țară cu multe limbi locale limba puterii colonialiste este opțiunea a doua în ordinea preferințelor pentru toată lumea.

- De ce standurile de înghețată sunt situate de cele mai multe ori unul lângă altul în mijlocul plajei, în condițiile în care, dacă ar fi răspândite pe toată plaja, ar fi mult mai bine pentru consumatori, iar pentru vânzători nu ar fi deloc mai rău? *Răspuns*: pentru că, la nivel individual, este rațional pentru fiecare vânzător să se deplaseze spre centru, indiferent ce face celălalt vânzător.

- De ce își exprimă unii oameni opțiunea de a vota chiar și atunci când aceasta nu poate produce niciun efect în privința rezultatului de la urne? *Răspuns*: pentru că votantul este victima gândirii magice sau acționează pe baza imperativului categoric.

- De ce indivizii de succes din societățile occidentale au, de regulă, o greutate mai mică decât restul oamenilor? *Răspuns*: pentru că ei cred (în mod corect) că ceilalți oameni întrețin (falsa) credință că oamenii cărora le lipsește autocontrolul într-un domeniu sunt mai predispuși la pierderea controlului, indiferent de situație.

- De ce oamenii se abțin de la tranzacții care ar fi în beneficiul tuturor, ca, de exemplu, faptul de a se abține să întrebe pe cineva aflat în față la coadă dacă nu ar fi dispus să își vândă locul? *Răspuns*: pentru că oamenii au o aversiune mai mare față de inegalitățile economice evidente decât pentru inegalitățile care nu se văd.

- De ce președintele Nixon a încercat să își creeze în fața sovieticilor o imagine de om înclinat spre un comportament irațional? *Răspuns*: pentru că

acest comportament urma să confere credibilitate amenințării de distrugere reciprocă, pe care sovieticii nu ar fi crezut-o altfel.

- De ce comandanții militari incendiază uneori pădurile aflate în subordinea lor (sau propriile nave)? *Răspuns*: pentru că se așteaptă ca adversarii lor să vadă că nu se vor mai putea retrage și, prin urmare, să nu intre într-o luptă costisitoare.

- De ce sunt chestiunile care țin de etichetă, în sine nesemnificative, atât de importante de cele mai multe ori? *Răspuns*: pentru că oamenii cred despre cei care se abat de la acestea că sunt nepăsători cu privire la părerea celorlalți despre ei.

- De ce oamenii lasă bacșiș taximetristului sau chelnerului chiar și atunci când se află într-un oraș pe care, probabil, nu îl vor mai vizita vreodată? *Răspuns*: pentru că îi doare să știe că ceilalți au o părere proastă despre ei.

- De ce investesc firmele în stocuri masive chiar și atunci când nu anticipează o întrerupere a producției? *Răspuns*: pentru că se așteaptă ca investiția lor să împiedice greva muncitorilor, asigurând astfel continuitatea de producție.

- De ce fiecare student dintr-un grup are impresia că ceilalți au înțeles mai bine un text obscur decât el însuși? *Răspuns*: pentru că fiecare dintre ei suferă de „sindromul fratelui mai mare", care îi face să creadă că ceilalți studenți nu cer ajutorul nu pentru că ar fi timizi, ci pentru că nu au nevoie de el.

- De ce, în multe adunări legislative, se votează prin apel nominal? *Răspuns*: pentru că susținătorii măsurilor populare folosesc această metodă pentru a-i descuraja pe aceia care, altminteri, ar vota împotriva lor.

- De ce sunt favorurile reciproce mai des întâlnite în adunările legislative ordinare decât în adunările constituante? *Răspuns*: pentru că schimbul de favoruri este susținut de un comportament de tip „serviciu contra serviciu", care cere o interacțiune continuă desfășurată pe o perioadă mai lungă de timp.

Cât de bine satisfac științele sociale calitative criteriile (1) – (4)? Dacă le luăm în ordine inversă, lucrările clasice nu sunt depășite. Mi-ar fi greu să iau în serios pe cineva care pretinde că literatura de specialitate clasică nu ne mai spune astăzi nimic pentru că descoperirile acesteia, în cazul în care sunt corecte, au fost pe deplin încorporate

în gândirea curentă. Scrierile acestea nu sunt de interes doar pentru anticari. Cu toate acestea, nu pretind că dialogul cu maeștrii din vechime este cel mai bun mod de a genera noi observații. De pildă, Thomas Schelling nu pare să fie, cel puțin nu într-un mod evident, dator unei anumite tradiții. Kenneth Arrow a redescoperit și generalizat o intuiție a lui Condorcet, dar nu a fost influențat de cel din urmă. Lucrările lui Daniel Kahneman și Amos Tversky nu s-au produs sub influența vreunui precursor, din câte știu. Am avut odată ocazia să îi indic lui Tversky că una dintre distincțiile pe care le face (între „efectul împroprietăririi" și „efectul contrastului") a fost anticipată de Montaigne și Hume. Acesta a răspuns că se simte bucuros să se afle într-o asemenea companie. Din moment ce oamenii de știință pe care i-am numit sunt responsabili pentru cele mai importante descoperiri în științele sociale din ultimii cincizeci de ani, devine evident că nu putem susține că singurul mod prin care se poate ajunge la intuiții noi este dialogul cu cei din vechime. Pentru alt exemplu, putem să îl luăm în considerare pe Tocqueville, care, după ce a decis să scrie *Ancien régime*, s-a abținut în mod deliberat de la lectura oricăror scrieri pe marginea subiectului, limitându-se la arhive și scrieri contemporane cu evenimentul studiat. Acesta a susținut că preferă să redescopere anumite adevăruri găsite și de alții, decât să fie influențat prea puternic de aceștia. Cu toate că se știe acum că a răsfoit și ce au scris alții pe marginea subiectului, capodopera sa exclude în cea mai mare parte influența acestora.

Acestea fiind spuse, dialogul cu trecutul poate fi foarte fructuos măcar pentru a identifica pozițiile care trebuie respinse. Este greu de imaginat că neomarxiști ca Weber sau Schumpeter ar fi putut scrie ce au scris dacă nu l-ar fi citit atât de atent pe Marx. Influența directă sau pozitivă este, de asemenea, foarte răspândită. Pare destul de probabil că anumite teorii recente cu privire la evoluția sistemelor de proprietate au fost influențate direct de David Hume, acesta fiind mai mult decât un simplu precursor. Lucrările lui Paul Veyne cu privire la psihologia tiraniei în Antichitate îi datorează multe analizei făcute de Hegel relației dintre stăpân și slugă. George Ainslie, care a contribuit puternic la transformarea intuițiilor elementare ale lui Freud în ceva convingător din punct de vedere analitic, probabil că nu ar fi ajuns la ideea sa dacă aceasta nu ar fi fost prezentă la Freud într-un mod incipient. Bănuiesc că lucrarea lui Bentham, *Political Tactics* (*Tactici politice*), este insuficient exploatată. În aceste cazuri, și în multe la care mă pot gândi, ideile inspirate de clasici trebuie să se poată susține singure odată ce au fost enunțate. Utilizarea corectă a clasicilor nu include folosirea apelului la autoritate în argumentare.

Științele sociale calitative, în cele mai multe instanțe pe care le-am menționat, îndeplinesc excelent criteriul (3). „Cotitura analitică" din științele sociale nu se bazează, din punctul meu de vedere, pe utilizarea unei metodologii cantitative, ci pe o preocupare aproape obsesivă pentru claritate și explicitare. (Preocuparea poate chiar lua o formă de-a dreptul obsesivă printre unii filosofi analitici.) Importanța operării

unei distincții între legăturile conceptuale și legăturile cauzale în rândul obiectelor de studiu câștigă tot mai mulți susținători. „Contextul" este din ce în ce mai mult interpretat ca „ceață", și nu ca „un mediu ferit". În comparație cu literatura mai veche, se merge mult mai puțin pe esențialism; cercetătorii se întreabă mult mai puțin ce sunt democrația sau socialismul „în sine". Se recunoaște că, deși sunt constrânse într-o anumită măsură de utilizarea lor, definițiile nu încearcă să surprindă esențe subiacente. Definițiile sunt stipulări arbitrare care trebuie judecate doar în funcție de cât de bine ne dau posibilitatea să dezvoltăm explicații corecte ale unor fenomene interesante.

Cred că există (sau, cel puțin, sper să existe) și o tendință de abandonarea a unei gândiri pe care o putem numi „abstractă nedeductivă". Pentru a fi onești, raționarea abstractă trebuie să fie supusă ea însăși ori disciplinei logicii deductive, ori unei referințe constante la fapte care demonstrează că aserțiunile abstracte au aplicații și *contribuie la sporirea cunoașterii*. În trecut, abstractizările trăiau în lumea lor. Sensul termenilor se putea schimba chiar în timpul argumentului, lăsând loc inferențelor nevalide. De pildă, Marx a „dedus" proprietatea privată din munca înstrăinată în felul următor: întrucât în capitalism produsul nu „aparține" muncitorului, în sensul că munca nu are sens, el trebuie să „aparțină" – să fie proprietatea – altcuiva, anume capitalistului. Un efect benefic al profesionalizării continue a științelor sociale în cele mai multe instituții de învățământ din Occident este acela că, în ziua de azi, este mult mai dificil să vii cu asemenea argumente. Recursul la analogie este mult mai rar întâlnit, cu toate că nu este absent. Cu toate că ideea „capitalului uman" reprezintă o extindere valoroasă a ideii de capital fizic, nu se poate spune același lucru despre „capitalul de consum" (Gary Becker), „capitalul cultural" (Pierre Bourdieu) și „capitalul social" (Robert Putnam). În cel mai bun caz, toate acestea sunt metafore inutile și benigne; în cel mai rău caz, acestea deschid nișe de cercetare nefructuoase și sugerează ipoteze cauzale false.

Este mult mai dificil să evaluăm în ce măsură științele sociale calitative satisfac criteriul (2). Printre istorici (și îi includ aici și pe aceia care scriu portrete de caracter ale unui regim), opiniile importante pierd uneori din importanță, pentru a o recăpăta ulterior. Ne putem gândi, de pildă, la ideea că Teroarea exista în stare latentă încă de la începutul Revoluției Franceze *versus* teza că fusese un rezultat al ascensiunii, care putea fi împiedicată, a lui Robespierre și al fugii regelui, care, de asemenea, putea fi împiedicată după ce delegitimase monarhia și aruncase țara în războaiele revoluționare. Conform celei de-a doua abordări, ne putem gândi că, dacă Mirabeau ar mai fi trăit, nu s-ar fi întâmplat nimic din toate acestea. Există întotdeauna istorici francezi care întreține una dintre cele două abordări, dar viziunea celor mai mulți este ciclică. Perspectiva abandonată azi poate fi dominantă mâine. Este greu să ne imaginăm vreo problemă istorică care să nu fi generat și să nu continue să producă fluctuații de acest

tip. Probabil că lucrurile vor sta astfel întotdeauna. Chiar dacă datele disponibile nu se pot modifica substanțial și pot fi trecute în revistă în totalitate, cum este cazul istoriei antice, generalizări noi produse de științele sociale neistorice pot confirma unele interpretări în defavoarea altora.

În științele neistorice există, în mod evident, progres, dar nu al cunoașterii în sens literal. Ne putem gândi iar la modul în care înțelegem de ce agenții raționali vor să treacă cu vederea unele dintre opțiunile pe care le au la dispoziție, incendiind podurile pe care le au în subordine, de pildă. Intuiția că un comportament care limitează poate fi un comportament *pe deplin* rațional împotriva unui adversar, și nu doar o protecție rațională împotriva unor înclinații iraționale, reprezintă un câștig ireversibil al înțelegerii. Cât de multe episoade ale incendierii unor poduri sau ambarcațiuni putem explica prin această idee este o chestiune de altă natură. Un comandant poate, ca în cazul lui Cortés, să incendieze navele doar pentru a împiedica trupele să fugă. Intuiția cu privire la posibilitatea unor majorități ciclice este ireversibilă, dar numărul de cazuri în care întâlnim așa ceva poate fi destul de mic. „Descoperirea" – care seamănă mai mult cu numirea unei intuiții comune preanalitice – că emoțiile au un timp de înjumătățire scurt nu poate fi ignorată, dar în fiecare caz pe care îl întâlnim ar putea exista anumite mecanisme care funcționează în alte direcții și care pot contrabalansa descompunerea spontană a emoției. Chiar dacă o explicație dominantă a unui eveniment sau episod este abandonată și apoi reînviată, cărămizile sau mecanismele care funcționează în abandonare și reînviere sunt aceleași. Inventarul sau dimensiunea instrumentarului nu se diminuează.

Științele sociale calitative nu satisfac foarte bine criteriul (1). După cum am remarcat deja, poate exista – de fapt, întotdeauna există – dezacord în cadrul aceleiași generații de istorici. Chiar dacă facem abstracție de doctrinele politice care îi determină pe marxiști și liberali să aibă perspective diferite asupra Terorii, tot rămâne destul loc pentru controversă între cercetători. De exemplu, atribuirea motivațiilor și credințelor actorilor din istorie este o chestiune foarte delicată. Un comportament pe care un istoric îl vede ca expresia unei stări mentale naive poate fi interpretat de altul ca un comportament strategic. Un aristocrat francez în exil în timpul Terorii poate spune că este de părere că monarhia va fi reinstaurată curând, dar, în același timp, poate face un contract pe termen lung pentru un apartament. Sunt slabe șanse ca atât contemporanii săi, cât și istoricii să îl creadă sincer în comparație cu nobilul care face contract după contract pe durate scurte, însă chiar și acesta din urmă ar putea să se comporte astfel doar pentru a da impresia unui optimism de nezdruncinat. Cu siguranță, dușmanii săi vor crede aceasta. Dacă istoricul nu este suficient de norocos să găsească un document al cărui scop nu pare a fi unul strategic, cum ar fi memoriile scrise de un actor istoric pentru sine, chestiunea nu poate fi decisă decât prin *judecată*. Și,

cu toate că un istoric bun este considerat astfel în mare parte pentru că face judecăți bune, acesta poate să și greșească.

Printre științele sociale calitative neistorice pot exista dezacorduri destul de substanțiale. Cei mai mulți economiști susțin uneori că multe dintre descoperirile economiei behavioriste care par să aducă dovezi în favoarea iraționalității se datorează faptului că sunt obținute prin experimente. Dintr-o varietate de motive, ce fac oamenii într-un laborator s-ar putea să nu corespundă comportamentului lor spontan de zi cu zi. Economiștii behavioriști au încercat să organizeze experimente pentru a combate această obiecție, mărind mizele, izolând subiecții de presiunea socială sau de ceilalți subiecți și de conducătorul experimentului și dând posibilitatea învățării în timp. S-a obiectat și că subiecții transferă în mod mecanic răspunsurile adaptive din afara laboratorului în cadrul experimentului, unde nu se pot adapta, utilizând, de pildă, strategii de tip serviciu contra serviciu în cazul unor experimente unice, în care un astfel de comportament nu are sens. Economiștii behavioriști au răspuns că nu este nimic mecanic în comportamentul subiecților în laborator, putându-se demonstra că aceștia sunt capabili de cele mai rafinate comportamente bazate pe stimulente. Dezbaterile continuă și este dificil să-ți faci o părere cu privire la ele ca nespecialist. Dacă am tendința să țin partea economiștilor behavioriști, de pildă, acest lucru se poate datora, într-o oarecare măsură, temeiurilor neintelectuale date de faptul că oponentul oponentului meu îmi este prieten.

Științele sociale cantitative

Mă voi apleca în cele ce urmează asupra științelor sociale cantitative. Acestea sunt de trei feluri: măsurarea, analiza de date și crearea de modele. Există legături între aceste tipuri de științe dat fiind că analiza de date este condiționată adesea de măsurare (pentru a determina datele) și de un anumit model (care să indice unde trebuie căutate datele). Mă voi concentra pe crearea de modele întrucât cunosc prea puține despre măsurare și analiza de date, dar voi spune, totuși, mai întâi, câte ceva despre acestea două.

Măsurarea unor entități precum consumul pe cap de locuitor, șomajul, ponderea corupției sau ce cred cetățenii de rând despre pedeapsa cu moartea este o operațiune dificilă în sine. Pentru ca măsurătorile consumului să poată fi folosite pentru comparații de-a lungul timpului sau între țări trebuie luat în considerare faptul că anumiți consumatori din perioade diferite de timp și din locuri diferite pot cumpăra obiecte diferite cu același venit, că veniturile lor sunt diferite și că bunurile disponibile pentru cumpărare pot fi diferite. Măsurarea șomajului este dificilă dacă economia subterană este foarte mare, dacă populația de studenți este mare, dacă există mulți oameni în penitenciare sau dacă există mulți oameni care au renunțat la a-și mai găsi un loc de muncă. Măsurarea corupției se bazează de obicei pe nivelurile percepute ale corupției,

așa cum sunt acestea determinate prin evaluări făcute de experți și prin sondaje de opinie. Din motive evidente, este dificilă obținerea unor probe independente prin care să se evalueze cât temei putem pune pe aceste surse. Chiar când converg, acest lucru s-ar putea datora unei părtiniri, pe care toate aceste surse o împărtășesc. Măsurarea opiniei publice presupune faptul că există un lucru stabil care poate fi măsurat. Însă se știe că modificări irelevante aduse întrebărilor pot determina diferențe majore la nivelul răspunsurilor. „Favorizați A?" poate stimula răspunsuri foarte diferite în comparație cu „Favorizați A dacă alternativa este B?", chiar dacă este foarte clar că B este singura alternativă la A. Pentru un exemplu, putem considera A, ca fiind aderarea unui stat la Uniunea Europeană, iar B non-aderarea. Aceste probleme pot afecta considerabil corectitudinea analizei de date.

Analiza de date (termen prin care mă refer la analiza statistică) este, într-un anumit sens, nucleul științelor sociale moderne. Dacă interesul nostru este să studiem fenomene sociale pe scară largă, știm că vom avea întotdeauna de-a face cu populații eterogene din mai multe puncte de vedere: sănătate, venit, statut familial, preferințe și reședință. Orice predicție deterministă conform căreia, odată expuși unui șoc extern X (de exemplu, o creștere în rata marginală a taxării), *toți* indivizii vor răspunde prin comportamentul B (de pildă, reducând nivelul muncii prestate) este sortită eșecului. Ce trebuie făcut este să determini cât de probabil este ca un individ dintr-o anumită categorie să răspundă într-un anumit mod, să stabilești o corelație între apartenența la acea categorie și răspuns. Aceste demersuri sunt adesea chestiuni foarte tehnice.

Pot fi și demersuri foarte periculoase, în sensul că prezintă unele capcane în care pot cădea chiar și cei mai abili cercetători. În capitolele 1 și 2 am menționat cinci astfel de capcane: extragerea de cunoștințe din date, ajustarea curbei empirice, arbitrarul în măsurarea variabilelor dependente și independente, problema distingerii cauzalității de simpla corelație și dificultatea de a identifica direcția în care se manifestă cauzalitatea. Practica obișnuită de a folosi valori „întârziate" ale variabilelor, corelând valoarea unei variabile la momentul t cu valoarea alteia la momentul $t + n$, duce la alte oportunități de improvizație. În plus, caracterul eterogen al unităților de analiză poate reprezenta o problemă. Într-o analiză statistică a cauzelor războaielor, nu este evident că Primul și al Doilea Război Mondial trebuie tratate la fel ca „războiul fotbalului" dintre Honduras și El Salvador. Însă este discutabilă și dorința de a elimina cazurile speciale pentru a îmbunătăți potrivirea curbei empirice pe o anumită ecuație. O altă problemă este aceea a selecției care poate apărea, de pildă, dacă sondajele de opinie se bazează pe interviuri prin telefon, care îi exclud pe aceia care nu au telefon sau care, din diverse motive, nu răspund la telefon.

Este imposibil pentru un nespecialist ca mine să îmi dau seama cât de des cad oamenii de știință în aceste capcane ale analizei de date. Standardul de referință trebuie să fie predicția sau „postdicția", care prezice o parte din observații din analiza

respectivelor date pe partea cealaltă. Simpla găsire a unei anumite logici în date nu convinge prea mult, date fiind posibilitățile infinite pentru manipularea datelor. Și acestea ar putea, în principiu, să fie reduse la un preangajament asumat de cercetători în mod public atât față de ipoteze, cât și față de proceduri, înainte să înceapă analiza propriu-zisă. Din câte știu, acest instrument este rareori folosit.

Crearea de modele este un exercițiu deductiv care începe cu presupozițiile și se încheie cu predicțiile. În științele sociale, modelarea prin alegere rațională este cea mai întâlnită formă, însă nu este nici pe departe singura. Cercetătorii propun tot mai mult modele evoluționiste, care nu presupun reacții raționale și nici măcar intenționate la evenimente. Cercetătorii au propus, de asemenea, modele mai mecaniciste ale comportamentului, adică modele fără fundamente clare sau justificări. Medicul englez Lewis Richardson, de pildă, a propus un model pentru cursa înarmării bazat pe două ecuații diferențiale liniare care nu aveau nicio argumentare clară în psihologie. Mă voi limita la folosirea alegerii raționale pentru a modela fenomenele sociale.

Modelarea prin alegerea rațională este foarte întâlnită în economie, unde se aplică mult peste domeniile tradiționale ale comportamentului consumatorului și producătorului. Există, de pildă, economia sinuciderii, economia prezenței la biserică și, mai general, analiza economică a oricărei activități care implică, într-un fel sau altul, o *alegere*. Tendința este aceea de a îndepărta bariera conceptuală care separă acțiunea intenționată de acțiunea rațională: dacă există o oportunitate pentru alegere, se presupune că ea va fi utilizată rațional. În științele politice, teoria alegerii raționale domină în mod evident datorită prestigiului mare pe care îl are, pentru că specialiștii sunt bine plătiți și pentru că, din ce în ce mai mult, devine un subiect obligatoriu de studiu în cadrul instituțiilor de elită (din Statele Unite). Aplicațiile se concentrează cel mai adesea pe politicieni (uneori și pe partide sau chiar pe state), mai puțin pe votanți și cetățeni. În sociologie, minoritatea care aplică alegerea rațională la temele tradiționale ale disciplinei a fost umbrită de economiștii care stăpânesc mai bine instrumentele (studiul încrederii este o excepție importantă). În antropologia socială, care (în Statele Unite) se deplasează în direcția științelor sociale slabe, teoria alegerii raționale nu este susținută de nimeni.

Este aceasta o știință? Să luăm în considerare criteriul (1). Putem presupune ca adevărat faptul că modelarea este o întreprindere corectă, în sensul că concluziile decurg din premise. Sunt aceste concluzii adevărate? Înainte de a trata acest subiect, vreau să fac observația că nu este clar dacă acestea au întotdeauna intenția de a fi adevărate, adică de a corespunde lumii reale. Mai degrabă, modelele reprezintă uneori o formă de literatură științifico-fantastică – o analiză a acțiunii și interacțiunii unor agenți raționali ideali care nu au existat niciodată în trecut și care nu vor exista niciodată în viitor. Analiza unor forme din ce în ce mai rafinate de strategii de echilibru, de pildă, nu este motivată de dorința de a explica și prezice comportamentul unor actori

individuali. Mai degrabă, motivația pare a fi una estetică. Doi dintre cei mai reputați teoreticieni ai echilibrelor, Reinhart Selten și Ariel Rubinstein, au spus răspicat că nu cred că modelele lor au ceva de spus despre lumea reală. Când vor să se pronunțe asupra unor aspecte ale lumii reale aceștia folosesc un tip de economie behavioristă sau de rațională limitată. Pentru a menționa un alt exemplu, teoria alegerii sociale – studiul axiomatic al mecanismelor de vot – ajunsese la un moment dat atât de matematizat și atât de evident irelevant pentru studiul politicii reale, încât una dintre cele mai importante reviste de economie, *Econometrica*, a impus un moratoriu asupra articolelor din acest domeniu.

O problemă interesantă în psihologia și sociologia științei este cât de mulți *practicanți secreți* ai abordării științifico-fantastice există în economie – cercetători care ascund fie față de sine, fie față de alții că aceasta este adevărata lor preocupare. Inventarea unor modele matematice ingenioase este o activitate bine plătită, dar, cu excepția unora ca Seltern și Rubinstein, plata nu va fi întreruptă decât dacă se poate susține că activitatea respectivă este relevantă; de aici și stimulentul pentru autoînșelare sau înșelăciune. A pune o astfel de întrebare poate părea în afara limitelor discursului academic, dar nu văd de ce. Dincolo de o anumită limită, normele academice de politețe ar trebui abandonate, argument susținut și de John Stuart Mill în *Despre libertate*, eseu care a devenit biblia libertății intelectuale încă de la apariție. Am renunțat deja la orice formă de politețe în ceea ce îi privește pe obscurantiștii care practică științele sociale slabe, nu văd de ce ar trebui să abordez *obscurantismul tare* în mod diferit.

Nu am dovezi directe nici cu privire la înșelăciune, nici cu privire la înșelarea de sine, prin urmare trebuie să abordez altfel chestiunea. Voi încerca să arăt că activitatea acelora care folosesc alegerea rațională este atât de departe de realitate, încât este greu să iei în serios faptul că tezele lor ar avea vreo legătură cu lumea. Cu riscul de a repeta ce am mai spus în capitolele anterioare, voi încerca să rezum argumentul meu în zece puncte.

1. Multe modele atribuie agenților motivații despre care se poate demonstra că aceștia nu le au. Un exemplu ar fi presupoziția, des întâlnită, că deprecierea timpului este exponențială.
2. Unele modele atribuie agenților motivații pentru care nu se oferă niciun fel de dovezi, cum ar fi o formă particulară a funcției de utilitate. Se poate, firește, stipula o funcție care să dovedească teoremele de posibilitate („un consumator rațional cu o utilitate marginală în scădere *ar putea* face X"), dar nu poți dovedi nimic despre lume. În multe situații, presupoziția că agenții sunt egoiști nu trebuie asertată pur și simplu, ci trebuie dovedită.

3. Cele mai multe modele ignoră progresele care s-au făcut în înțelegerea alegerii, asociată cu teoria perspectivei. Din moment ce teoria este enunțată în termeni de câștiguri și pierderi dintr-un anumit punct de referință, nu analizează comportamentul în funcție de maximizarea unei funcții obiective definită de rezultate. Formatul maximizării poate fi indispensabil pentru manipularea matematică, dar este irelevant pentru o mare parte a comportamentului real.
4. Unele modele atribuie agentului capacități cognitive pe care acesta le poate avea, dar nu le practică. De pildă, un agent poate să nu realizeze în mod spontan inducția inversă necesară în multe aplicații.
5. Multe modele atribuie agenților capacități cognitive despre care se poate demonstra că nu le au. Argumentul este atât de banal, încât situația este aproape jenantă: cum poate un economist să presupună că un agent are abilitatea să efectueze calcule pe care economistul le poate prezenta în câteva pagini? Este foarte puternică tentația de a spune „Ei hai! Fii serios!".
6. Multe modele îi atribuie intenții pe baza intereselor obiective ale agentului în situația dată, fără să se întrebe în vreun moment dacă alte motivații – emoțiile sau normele sociale – nu pot și ele să aibă un rol. Aproape că ne-am face de râs dacă am încerca să explicăm comportamentul în situații turbulente și conflictuale precum revoluțiile, de pildă, atribuind un motiv legat de interesul personal pe termen lung al agentului (redus la valoarea sa prezentă prin depreciere exponențială). Alte modele atribuie intenții pe baza unor rezultate curente, neglijând astfel atât posibilitatea ca unele calcule să fie greșite, cât și pe aceea ca altele să nu fie mai mult de simple pariuri.
7. Multe modele nu țin cont de ceața incertitudinii care învăluie cele mai importante decizii, în special când consecințele relevante se află în viitorul îndepărtat. Încercările de a depăși această problemă atribuind probabilități subiective rezultatelor sunt, de cele mai multe ori, arbitrare, ca în cazul în care cercetătorii recurg la principiul rațiunii insuficiente pentru a manipula o distribuție uniformă de probabilitate.
8. Unele modele stipulează colectivități (clase sau state) ca și cum ar fi agenți unitari, fără să pună în discuție (în cazul claselor) problema actelor parazitare sau (în cazul statelor) structura deciziei colective.
9. Unele modele presupun că deviațiile de la raționalitate sunt fie (1) temporare, fie (2) se vor anula ca agregat. Cu privire la (1), se poate spune că deviațiile temporare pot avea rezultate durabile dacă situația are o structură de tip „capcană de homari". Cu privire la (2), economia behavioristă a

arătat că, din moment ce multe deviații sunt sistematice, și nu întâmplătoare, nu avem vreun motiv să ne așteptăm că acestea vor dispărea când vom considera lucrurile la nivel de agregat. „Paradoxul bonusului de acțiuni" este un astfel de exemplu.

10. Multe modele se mulțumesc cu faptul de a explica comportamentul argumentând că este punctul de echilibru al unui joc, fără să facă pasul următor și să arate de ce, în situații cu echilibre multiple, s-a realizat acest echilibru particular. Mai mult, acestea se opresc rareori pentru a se întreba dacă nu cumva lumea se schimbă atât de rapid încât nu este timp suficient pentru stabilirea unui echilibru.

Am remarcat câteva răspunsuri tipice la aceste obiecții. Un răspuns este acela de a cita problema hermeneutică a stabilirii de motivații într-o manieră necirculară. Sunt de acord că problema aceasta este spinoasă, dar nu insurmontabilă, și, chiar dacă ar fi, tot nu ar fi suficientă pentru a justifica atribuirea gratuită de motivații. Un alt răspuns este acela de a propune înlocuirea raționalității reale (sau a agenților vii, reali) cu „raționalitatea de tip ca și cum". Am arătat că aceasta este o stratagemă care poate funcționa doar dacă sunt satisfăcute două condiții. Pe de o parte, am fi în măsură să indicăm un mecanism capabil să imite raționalitatea, *chiar până la* nivelul anexelor matematice. Problema e că nimeni nu a reușit să propună un astfel de mecanism. Recurgerea la selecție nu ne ajută prea mult din moment ce există o imensă discrepanță între caracterul brut și imperfect al mecanismelor de selecție socială și procesele de raționare ultrasofisticate atribuite agenților. Pe de altă parte, am fi în măsură să indicăm predicții atât de precise, încât am fi constrânși să acceptăm teoria chiar dacă nu înțelegem modul în care funcționează. Problema e că științele sociale nu fac predicții care să se apropie nici măcar vag de un astfel de statut.

De fapt, dacă ne uităm mai atent la satisfacerea criteriului (1), vom observa coexistența în timp a unor teorii diferite cu explicații și predicții foarte diferite. În special macroeconomia se evidențiază din acest punct de vedere. Nu cunosc statutul curent al keynesianismului, monetarismului și al teoriei așteptărilor raționale, dar cred că aș fi remarcat dacă vreuna din ele le-ar fi învins pe celelalte oferind pentru o perioadă de timp un flux stabil de predicții foarte exacte. Este oare un adevăr bine stabilit că avem nevoie de o bancă centrală independentă sau este doar o modă? De decenii, economiștii propun explicații radical diferite cu privire la rata mare a șomajului în Europa, fără să fie în stare să ajungă la un consens. Oare oamenii muncesc într-adevăr mai puțin dacă suportă taxe mai mari? Opiniile experților sunt diferite. Acestea nu sunt chestiuni care țin de zonele de periferie ale domeniului de cercetare, unde vor exista întotdeauna dezacorduri, ci sunt probleme fundamentale, care formează nucleul dur al disciplinei.

Nici criteriul (2) nu este satisfăcut, după cum sugerează prefixul „neo-", ca în neokeynesianism sau neofuncționalism. Criteriile (3) și (4) sunt satisfăcute cu ușurință; de fapt, sunt suprasatisfăcute. Profesionalizarea științelor sociale în Statele Unite, la care m-am referit mai devreme, are, ca revers al medaliei, un caracter îngust, care sufocă, și o obsesie dăunătoare sieși pentru clasificarea departamentelor în cutare sau cutare categorie. De fapt, cred că faptul de a fi cotat bine este o stare care constituie esențialmente un produs derivat al căutării cunoașterii, și nu al căutării unei anumite poziții într-o ierarhie. (După cum am văzut, Proust face o remarcă asemănătoare cu privire la muzicieni.) Mai mult, pentru că dialogul cu trecutul este pus în sarcina unei subdiscipline mai puțin prestigioase, anume istoria gândirii economice, sau în sarcina mai prestigioasei istorii a gândirii politice, caracterul îngust se perpetuează. Însă, deși patologiile obscurantismului tare sunt preferabile acelora ale obscurantismului slab, aceasta nu înseamnă că trebuie să îl și preferăm.

Viitorul științelor sociale

Oare ne putem descurca mai bine? Pot aspirațiile științifice ale științelor sociale – căutarea predicției, a preciziei și a caracterului definitiv – să fie satisfăcute la un anumit moment în viitor? Istoria științei ne învață să fim foarte precauți atunci când facem afirmații despre ce poate să facă o anumită știință și ce nu poate să atingă. „Toată lumea a râs" când Descartes a spus că animalele sunt mașini, dar cine râde la urmă? Orice încercare de a aborda această problemă trebuie să fie atât provizorie, cât și speculativă.

Asimilarea descoperirilor din neuroștiințe va furniza o bază solidă pentru psihologie și va rezolva unele controverse curente. De pildă, deja s-a susținut că scanările creierului sprijină deprecierea cvasihiperbolică, și nu pe aceea hiperbolică. Am putea chiar bate în cuie distincția dintre mânie și indignare dacă s-ar demonstra că aceste două emoții activează centre diferite în creier. Neurofiziologia dependenței va continua să sporească ceea ce știm despre acest comportament frapant și autodistructiv. Însă trebuie remarcat că toate aceste aplicații presupun o *motivație* pentru acțiune. Sunt numeroase și aplicațiile asupra *percepției*, precum fenomenele de completare a golurilor. Spre deosebire de acestea, nu există încă și nici nu putem întrevedea că ar apărea în curând o neuroștiință a *credințelor* propoziționale (și ale interacțiunilor motivate de credințe). Să luăm în considerare un caz trivial de actualizare. Mă gândeam că s-ar putea să plouă mâine, dar, întrucât cerul s-a înseninat, mi-am revizuit predicția și, prin urmare, mi-am schimbat planurile. Nu cred că doar lipsa mea de imaginație (sau de expertiză științifică) mă determină să am dubii că vom fi curând în măsură să identificăm neurofiziologia chiar și unui proces atât de simplu; desigur, multe procese de formare a credințelor sunt cu mult mai complexe. Dacă am dreptate, atunci jumătate din modelul credință-dorință al acțiunii va rămâne pentru multă vreme de acum

înainte inaccesibil unei abordări științifice. Același lucru este chiar mai valabil dacă depășim modelul și ne punem întrebări cu privire la „neuronii inspirației", aceia care produc opere de artă.

Există două tipuri de motive pentru care științele sociale nu reușesc să facă predicții sau să producă explicații în sensul tare al termenului. Un motiv este că până și pentru credințe și preferințe date acțiunea poate fi în bună măsură indeterminată (adică, imprevizibilă). În cazul deciziei în condiții de incertitudine sau în cazul deciziilor care au un grad mare de complexitate, comportamentul poate fi influențat de ceea ce Keynes numește „spirit animalic", și nu de trăsături identificabile ale situației la care reacționăm în moduri identificabile. Este adevărat că oamenii acționează și după reguli de tip empiric în aceste cazuri: problema e că acestea sunt mult prea numeroase. Pot exista, de asemenea, puncte focale competitoare: *fă cum s-a făcut până acum* versus *fă ce face vecinul*. Nu spunem prea mult prin ideea că oamenii urmează „satisfacerea" (*satisficing*), și nu „maximizarea", dacă nu putem identifica *ex ante* ce reprezintă un nivel satisfăcător.

Un alt motiv este slaba noastră înțelegere a mecanismelor de formare a preferințelor. Indivizii sunt supuși unor înclinații competitoare, a căror tărie relativă, în orice situație dată, este adesea indeterminată. Dacă te amenință, te vei înfuria sau ți se va face frică? Dacă apar ambele efecte, care dintre ele va domina? Dacă fac o pauză din viața mea profesională agitată, timpul liber va fi la fel de agitat sau, dimpotrivă, destins în totalitate? Dacă țara mea trece de la dictatură la democrație, oare eliberarea mea de autoritatea politică mă va face să abandonez autoritatea religioasă sau, dimpotrivă, mă va face să o caut? Am susținut că suntem adesea incapabili să răspundem unor astfel de întrebări cu privire la formarea preferințelor *ex ante*, cu toate că, după ce faptul s-a consumat, putem fi în măsură să identificăm mecanismul dominant.

Am putea oare diminua această indeterminare identificând fie *condițiile de declanșare*, fie *probabilitățile de declanșare*? Am discutat prima posibilitate în capitolul 2, ajungând la concluzii destul de sceptice, prin urmare mă voi referi aici la cea de-a doua posibilitate. Am putea, de exemplu, să specificăm o distribuție în populația nivelurilor de satisfacere care, cel puțin, ne-ar îngădui să prezicem comportamentul agregat? Până la urmă, acesta este în mare obiectul de activitate al științelor sociale. Asemănător, dacă presupunem că o anumită parte a populației se va teme sau se va înfuria în urma unei amenințări ori că fiecare dintre noi ar putea reacționa în oricare dintre cele două moduri, atunci nu ar fi posibil să cuantificăm aceste probabilități?

Să presupunem că am putea, dar tot nu cred că am câștiga prea multă putere explicativă din cauza importanței imense a *contextului*. Mai devreme am spus că progresul în știință decurge adesea în urma abstragerii din context. Acest lucru este valabil și pentru științele sociale, în sensul că nu putem identifica înclinații sau mecanisme decât dacă ignorăm foarte multe trăsături ale situației și creăm un cadru din care acestea

lipsesc. Noțiunea intuitivă de conformism, de exemplu, este complexă. Aceasta include încrederea în ceilalți ca sursă de informare, dorința de a fi precum ceilalți, dorința de a nu ne asuma riscuri și dorința de a nu le lăsa celorlalți o impresie proastă. Dacă pentru un scop anume al cercetării vrem să utilizăm (să zicem) dorința de a fi precum ceilalți ca definiție a conformismului, atunci trebuie să o izolăm de celelalte trei dimensiuni ale conformismului prin crearea unui cadru experimental (sau identificarea unuia real) în care celelalte elemente nu sunt prezente. Probabil că vom descoperi că oamenii vor, într-adevăr, să fie precum ceilalți, în timp ce alții vor să se distingă de ceilalți. Am putea chiar specifica proporțiile sau probabilitățile. Însă într-o situație reală celelalte dimensiuni ale conformismului vor fi și ele prezente și ar putea produce un efect contrar. Faptul de a ști în ce mod sunt distribuiți oamenii în funcție de patru dimensiuni ale conformismului s-ar putea să nu ne ajute prea mult în prezicerea *comportamentul* conformist. Analiza, desfacerea lucrurilor în părți componente, este fezabilă. Sinteza, asamblarea elementelor, este mult mai grea, poate chiar imposibilă.

Ne putem imagina ușor cum ar putea răspunde acestei situații un cercetător hotărât să dovedească posibilitatea unei *științe* sociale, dar nu voi continua cu discuția. Unii dintre noi sunt impresionați și copleșiți de complexitatea și instabilitatea comportamentului uman. Alții au o credință intuitivă într-o regularitate subiacentă, care, odată scoasă la lumină, ne va permite să punem științele sociale pe picior de egalitate cu științele naturii – fizica, chimia, geologia sau meteorologia. Viitorul va decide, însă eu am încercat să arăt cât se poate de clar pe ce aș paria și de ce.

BCU IASI/CENTRALA UNIVERSITARA

BCU IASI/CENTRAL UNIVERSITY LIBRARY

INDEX

11 septembrie 2001, 64
  vezi şi World Trade Center, atacul de la
Abu Sayyaf, 68
abuzarea copiilor, 283
acţiune
  colectivă, 48, 363, 365, 369-371, 373, 394
  interpretare a., 62
  modelul credinţă-dorinţă, 75, 429
  raţională, 190
  vezi şi teoria alegerii raţionale
acţiune din umbră, 200
adaptarea hedonică, 66
ADN, codoni ai, 264, 279
adulter, 275-276
adunări constituante, 19, 336, 378, 380, 384, 392, 394, 403, 419
aeronava *Concorde*, 17, 417
aeroportul din Oslo, 387, 394
afront, 155-156, 256
Agnew, Spiro, 399
agregare
  a credinţelor, 382
  a preferinţelor, 374, 382, 384, 388, 396
Ainslie, George, 420
ajustarea curbei empirice, 424
Albertine, din *A la recherche du temps perdu*, 251
alcool, 47
alcoolici, 51, 156
alegere, 114-115, 117, 123, 133, 154, 161-162, 164, 173, 187, 189-193, 199, 202-205, 208, 213, 218, 225-226, 228, 230, 236, 238, 260, 262-263, 267, 270, 274, 278, 281, 283, 287-288, 291, 297, 302, 305, 307-310, 319-320, 323-324, 329, 365, 373-376, 378, 380, 382, 384, 405, 425-426
Alegerea Sofiei, 89
alfabetizare, 377-378
alienare, 289
amae, 159
amânare, 87, 115, 219, 228, 230-231, 233, 244-245
ambiguitate, aversiune faţă de, 90
ameninţări, 276, 280, 311-312, 375, 389-392, 402
amigdala, 251-253
aminoacid, 264, 266-267
amoralism familial, 104
analogie, 146, 226, 247, 249, 261, 421
Andersen, Hans Christian, 12, 287, 296, 351, 354-355
Andrewes, Episcopul Lancelot, 77, 81
anorexie, 64
anosognozie, 258-259
antisemitism, 184
a ochi în faţa ţintei, 115
apreciere, 79, 148
aprobativitate, 102-104, 107, 112
argumentare, 134, 374-380, 393-394, 398, 412, 420, 425
Aristodemos, 92
Aristotel, 12, 123, 149, 241, 382-383
  vezi şi indignare aristotelică
aroganţă, 162, 175, 221, 413
Aron, Raymond, 37
Arrow, Kenneth, 59, 388, 420
aşteptare, 154, 197, 252, 278, 316-317
Atena, 94, 384
atribuire, eroarea fundamentală de, 183-184
Austen, Jane, 244-245
autokinetic, efect, 350
autocontrol, 98, 232, 271, 324, 418
autopaternalism colectiv, 234
autosugestie, 79

BCU "LUCIAN BLAGA" CENTRAL UNIVERSITY LIBRARY

aversiune față de risc, 391
avocatul diavolului, 33
avort, 94-95, 136, 191, 234, 274, 278
axioma independenței, 192, 210

bacșiș, 19
vezi normele privitoare la bacșiș
Bambara (Mali), 205
bârfa, 111
Barnave, Antoine Pierre Joseph Marie, 391
Bayes, Thomas, 197-198
Bătălia Sexelor, 303, 306-308
Becker, Gary, 421
Belgia (după război), 231
Bemba, Rhodesia de Nord, 339
beneficii procesuale, 371
Bentham, Jeremy, 379, 384, 394, 420
Bergerac, Cyrano de, 146
Bernard, Claude, 411
Bertram, Edmund, în *Mansfield Park*, 244
bicameralism, 69, 234-235, 381-382, 391, 394
Blestemul învingătorului, 213, 218
Blum, Léon, 283
Bogart, Humphrey, 328
Bogota, 108
Boswell, James, 34, 208, 227
Bourbon, Antoine de, 68
Bourdieu, Pierre, 421
Brandeis, Louis, 379
Bruun, Ketil, 59
bucurie, 16, 92, 150, 242, 415
budism, 155
bulgărele de zăpadă, 354, 363, 365
Bulstrode, în *Middlemarch*, 242
bunul samaritean, 161, 163
Burgundia, în *Regele Lear*, 96
Burke, Edmund, 408-409
Burroughs, E.R., 313

calmare, 220
Calvin, Jean, 65
calvinism, 16, 65, 133
vezi și predestinare
Campionatul Mondial de Fotbal (1998, 2002, și 2006), 16, 83, 151
cancer
intestinal, 200
al pancreasului, 36-37
pulmonar, 44
stadii terminale, 212
capcană de homari, 427
capitalism, contradicția c., 291
Carraciolo, Antoine, Episcop, 68
carteluri, 369, 372
Carter, Jimmy, 23
cascade informaționale, 359
câștigătorul ia tot, 94
cauzalitate
corelație, 351
direcția, 424
lanț cauzal, 24, 40, 43, 45
socială, 285
vezi și efectul interacțiunii
căsătorie-legământ, 155, 233
certitudine, 36
Chasteller, doamna de, în *Lucien Leuwen*, 232, 240-241
Châtillon, Cardinal de, 68
Cherokee, 311
China
doar un copil, 337
reforma economică în, 311
Chirac, Jacques, 314, 416
ciclul „pânzei de păianjen", 293-294
Claudius, în *Hamlet*, 236-237
cleptocrația, 399
cluburile de Crăciun, 217
cocaină, 231
coincidență, 58
colegii electorale, 401
Coleridge, Samuel, 233, 235
comedie, 35
compensații, 77, 83, 110, 120, 139, 159, 255
completare a golurilor, 429
comportament
altruist, 37, 255
excesiv, 230, 232
externalități ale, 289

BUCHAREST CENTRAL UNIVERSITY LIBRARY

fobic, 64
inteligibil, dar irațional, 63
comunism, 37, 126, 137, 352, 413
Concursul de Frumusețe, 319
condiții de declanșare, 430
Condorcet, Marie Jean Antoine, 383, 394
conformism, 176, 338, 348-349, 354, 431
consolidare
învățarea prin c., 16
plan de c., 261
constituție, 307, 381, 397, 404-407
constrângeri asupra comportamentului, 248
consum de alcool în campus, 353
conștiință, 65, 67, 86-87, 92, 236, 371
conștiință colectivă, 332
control al pulsiunilor, 87
control judiciar, 397, 407
Convenția Federală (Philadelphia, 1787), 70, 85, 379, 390, 406
convenții, 243, 300, 312, 335-336, 347, 391
cooperare, 106, 108, 133, 143, 176, 219, 271-272, 301-302, 305-307, 309, 320-321, 323, 329, 330-331, 346, 366-367, 369-372, 417
coordonare, 297, 300, 302, 307-309, 356, 363
Cordelia, în *Regele Lear*, 96
corelație, 30, 35, 41, 43-45, 57-58, 60, 255, 351, 424
Corsica, răzbunarea în, 339
Cortés, Hernán, 313, 422
corvée, 408
credibilitate, 33, 45, 71, 143, 242, 311-312, 331, 389-391
vezi și promisiuni
credințe contramotivaționale, 136
Crew, Frederick, 412
cristalizare, teoria lui Stendhal, 156
criza rachetelor cubaneze, 306
criza de apă, 108
Cronica unei morți anunțate (García Márquez), 239
culegerea de informații, 187, 199-200, 202, 383
cumpărături prin comparație, 207
curba Phillips, 295
cursa înarmării, 304, 309, 425
Curtea Supremă
din Israel, 398
din Norvegia, 397
din Statele Unite, 406
custodia copilului, 17, 206, 416
cușer, reguli, 89
cvasi-credințe, 78
cvasi-emoții, 78

Darwin, Charles, 129, 277-279, 411
de Gaulle, Charles, 176, 231, 369, 406, 410
decizia
colectivă, 7
costurile, 207
încâlcirea firelor, 63
scurtcircuitarea mașinăriei, 63
trunchierea mașinăriei, 63
decizii de titularizare, 375
Declarația Drepturilor Omului și Cetățeanului (1789), 398
deconstructivism, 412
demobilizare, 388, 394
democrație și oligarhie, 373, 383
Denton, Jeremiah Andrew, 211
dependența de semnale, 231
dependență de cale, 97, 360
dependență
recidivă, 123
și sevraj, 292
depreciere
cvasihiperbolică, 429
exponențială, 119, 121, 426-427
hiperbolică, 121, 417, 429
pastilă a, 204
pură a timpului, 118
a timpului, 426
depresie, 276-277, 284
Derrida, Jacques, 412
Descartes, René, 73, 106, 148, 248, 251, 429
Despre democrație în America (Tocqueville), 310, 351, 413-414
Despre iubire (Stendhal), 145
Despre mânie (Seneca), 232

BCU CENTRAL UNIVERSITY LIBRARY

destrămare, 354, 363-365, 372
dezamăgire, 150
dezarmare unilaterală, 368
dezgust, 158
Diagoras, 131
dializă, 131
Dilema Călătorului, 319, 321, 323
dilema hermeneutică, 67
Dilema Prizonierului
  iterată, 369
  cu mai multe persoane, 364
Dion Hrisostomos, 88, 99
disonanță cognitivă
  producerea, 63
  reducerea, 63, 79, 135, 353
dispreț, 148, 152-153, 155, 170, 175, 222, 275, 295, 334-335, 339, 345
distincția dintre a vrea și a dori, 90-91
disulfiram, 205-206, 231
divorț, 35, 233-234, 391
donații în scopuri caritabile, 102
Donne, John, 50, 75, 197, 323, 326
dorință
  arzătoare, 50
  pătimașă, 146
Dostoievski, Fiodor, 178, 240
DP, 302, 304-305, 309, 312, 319-321, 325, 329, 364-365, 367, 369
  vezi Dilema Prizonierului
dramă, 238, 242
drept de vot, 382-383
dreptul penal islamic, 408
droguri, 57, 123, 166, 178, 189, 286
dueluri, 213, 339, 341
Duport, Adrien-Jean-François, 391
durere, 28-30, 34, 36, 142, 145, 147, 158, 181, 204, 212, 228, 256, 277, 317, 415
Durkheim, Emile, 247

echilibre de coordonare, 307
echitate, 48, 96, 105, 107-108, 299, 316
economia de piață, 396-397
economie comportamentală, 7
educație, 24, 56, 62, 66, 69, 73, 85, 94, 104, 118, 170, 295-296
efect de domino intern, 87, 225
efect de întărire, 288
efecte dependente de frecvență, 35
efectul certitudinii, 127, 192, 213, 217, 223
efectul de compensare, 48, 53
efectul de disjuncție, 213, 219
efectul Giffen, 46
efectul de halo, 179
efectul interacțiunii, 56, 107
efectul împroprietăririi, 287, 420
efectul Othello, 156
efectul de răspândire, 48,53, 176
efectul Veblen, 46
eficiență, lipsă de considerație pentru e., 213
egoism, 101, 105, 119, 189, 203, 324
Elliott, Anne, în *Persuasiune*, 145, 147
emigranți, aristocrați francezi e. în Londra, 71
emoție
  și acțiune, 152
  antecedente cognitive ale, 147, 251
  pot afecta credințele, 221
  evaluative, 148
  și evoluția, 275
  și expresii fiziologice, 147
  și obiecte intenționale, 147
  răzbunării, 417
  și stimulare fiziologică, 147
  surprindere, 158
  timp de înjumătățire al, 414, 417, 422
  și transmutație, 157
  valență, 147-148, 151
  vezi și apreciere; aroganță; bucurie; calmare; cvasi-emoții; dezamăgire; dezgust; dispreț; durere; euforie; frică; gelozie; indignare; invidie; iubire; jubilare; milă; mândrie; mânie; ranchiună; regret; rușine; speranță; tristețe; ură; vină
empatie, 61
Erklären, 61, 73
eroare
  Concorde, 158, 212
  a costului investit, 211, 221
  a jucătorului de ruletă, 132, 211
  vietnameză, 158

BCU IAȘI - CENTRAL UNIVERSITY LIBRARY

eroarea fundamentală de atribuire, 183-184
erori motivate („fierbinți"), 134
erori nemotivate („reci"), 134
esențialism, 184, 421
etica iudaică, 99
eu, 86
euforie, 57, 93, 146, 151
Euripide, 122, 239
euristică, 214, 217
euristica disponibilității, 132, 217, 414
euristica reprezentativității, 217, 414
evoluție, 55, 264-265, 270, 345
evrei
  salvarea, 179
  vezi și antisemitism
execuții extra-judiciare, 398
experimente de comparare a liniilor, 360
experimentele cu lumina în mișcare, 360
experți, 59, 61, 66, 69, 70, 79, 92, 130-131, 405, 424, 428
explanandum, 260, 414
explanans, 21-22, 27-28
explicație
  prin alegerea rațională, 187, 205, 208
  prin consecințe, 27, 260, 446
  și corelație, 43
  evoluționistă, 276, 284
  funcțională, 19, 258, 340-341
  intențională, 21
  și interpretare, 61
  și istorisirea, 38
  și întrebările de forma „de ce?", 414-419
  și necesitare, 43
  și predicție, 40
  statistică, 309
  structurală, 37
externalități, 289, 291-292, 295, 302, 337-338, 346, 365-366, 369
  vezi și internalități
extragerea de cunoștințe din date, 424
extremism, 56

Fabius Cunctator, 63
fals efect al consensului, 133
fapt „nou", 61
FARC (Forțele Armate Revoluționare din Columbia), 68
Fatah, 68
Federalistul (Madison), 169
Fedra, 239
fenotip, 264, 269
Ferguson, Adam, 289
fericire, 85, 92, 128, 145, 147, 149, 150-151, 158, 225, 258
Fericitul Augustin, 111, 122
Festinger, Leon, 32, 42
fond de grevă, 312, 390
formarea constantă a numărului (de telefon), 262
formarea credințelor, 131, 144, 154, 156
  vezi și autoînșelare; cascade informaționale; gândirea antideziterativă; gândire magică; gândire deziterativă; ignoranță pluralistă; inferență, eroare statistică; zvonuri
Forțele Armate Revoluționare din Columbia, *vezi* FARC
fotbal, 282, 401-402, 415, 424
Franța
  din *Ancien Régime*, 335, 355, 407, 413-414
  Consiliul Constituțional, 397
  insurecția muncitorilor din Paris (1848), 355
  invadarea de către Germania în 1940, 200
  războaiele religioase din (1562-1598), 68
  Republica, 309
  Restaurarea, 355
  Revoluția, 295
  vezi și adunările constituante
Freud, Sigmund, 79-81, 86, 99, 114, 138, 252, 275, 420
frică
  cercetări în neurofiziologie cu privire la f. (în cazul șobolanilor), 251
  a crede ceva care provoacă f., 156
  viscerală *versus* de prevedere, 84
Friedman, Milton, 38, 42
fructul oprit, 48

BCU CENTRAL UNIVERSITY LIBRARY

fumat, 118, 122, 133, 138-139, 187-188, 195, 203, 230-231, 338
Fundația Carnegie, 101

gândire antideziterativă, 157, 221, 358-359
  deziterativă, 157, 194, 199, 214, 217, 358-359
  magică, 323, 415
gânduri contrafactuale, 150
Geisteswissenschaften, 61
gelozie, 71, 251, 275-276
gene, 44, 270-271, 275-277, 279, 417
genetică, 247
genotip, 264
Genovese, Kitty, 17, 26, 42, 49, 162, 180-181, 186, 306, 352, 416
Germania nazistă, 278
glorie, 65, 92-94
Goldwyn, Sam, 345
greutăți și măsuri, 309
Guise, Henri, Duce de, 68
Gutman, Kasper, 328
guvern
  și controlul judiciar, 406
  înfometarea opoziției, 405
  ignorarea restricțiilor (impuse asupra acțiunilor), 406
  și justiție politică, 404
  și legislație în interes propriu, 405
  și majoritatea populară, 406
  manipularea fluxului de informație, 404
  manipularea sistemului electoral, 404
  manipularea politicii monetare, 404
  ocolirea restricțiilor (impuse asupra acțiunilor), 405

hainele cele noi ale împăratului, 351
Hamas, articolul 22 din Carta H., 128
Hamilton, Alexander, 86
Hamlet, 17, 79, 208, 236-237, 244-245, 417
Hamsun, Knut, 178-179, 185, 240
Hegel, Georg Wilhelm Friedrich, 289, 403, 420
Heisenberg, Werner, 249
Henric al IV-lea, 68
  vezi și Navarre, Henri de
hermeneutica suspiciunii, 111
heroină, 231
Hicks, John, 171
hiperraționalitate, 207, 258
hipersomnie, 276
Hirschman, Albert, 128
Hitler, Adolf, 66, 206
Hobbes, Thomas, 302, 397
Homer, 232
Hume, David, 23, 102, 402-403, 420

Iago, în *Othello*, 157
iertare, 65, 159, 327
ignoranță
  motivată, 118
  pluralistă, 285, 304, 351-353, 355
  văl al i., 105
imigrație, 171
imitație, 280-281, 339
imparțialitate, 85, 256, 403
imperativul categoric, 214, 219-220
Imperiul Otoman, 408
imunitate, 391
inabilitatea de a proiecta, 218
incendierea podurilor sau a navelor, 313
incertitudine, principiul i. (Heisenberg), 249
inconștiente
  mecanisme, 79
  stări mentale, 79
indeterminare, 384, 430
indiferență, curbe de i., 165-166
indignare
  aristotelică, 149, 157
  carteziană, 152
individualism metodologic, 27, 47, 52, 59
inducție inversă, 317, 322
inferență statistică, erori ale i.s., 131
inovație, 281
insomnie, 276
instituții, încrederea în i., 330
Insula misterioasă (Verne), 243
insurgenți ceceni, 68

BUCHAREST UNIVERSITY LIBRARY

inteligibilitate
absolută, 239
și raționalitate, 62
relativă globală, 239
relativă locală, 239
intenție, 268, 290, 334
intenționalitate, niveluri ale i., 278
interacțiune
repetată, 369
socială, 286, 302
unică, 213
interes, 364, 373, 377-380, 392-393, 398-399, 401, 403-406, 420, 424, 427
internalități, 291-292, 296
vezi și externalități
interpretare
a acțiunii, 62
prin consecințe, 236
a operelor de artă, 236
Intifada a doua, 66
intoxicare, 85
intuiții cauzale, 40
invidie, 316, 340, 414
ipocrizie
cultura i., 352
forța civilizatoare a i., 377, 403
iraționalitate auctorială, 237
iubire romantică, 159
izbăvire, 92

încălzire globală, 268, 285
încredere
care determină un caracter demn de încredere, 329
motive pentru, 326
oarbă, 326, 330
înlocuirea unei cauze, 36
întrebări de forma „de ce?", 414-419
înțelegere, 391, 398, 427, 429-430
învățare bayesiană, 197, 350
învinovățirea victimei, 259

Jackson, Robert, 407
Jante, Legea lui, 338
Jaurès, Jean, 283
Jocul Asigurării, 303
Jocul Dictatorului, 315
Jocul Încrederii, 316, 389
Jocul Lașului, 306-307
Jocul Miriapodului, 319-321, 323
Jocul Punctului Focal, 302-303, 308
Jocul Telefonului, 302-303, 307
Jocul Ultimatumului, 315-318, 320, 335
jocuri
experimentale, 107
secvențiale, 310, 317, 375
vezi și Bătălia Sexelor; Concursul de Frumusețe; Dilema Călătorului; Dilema Prizonierului; Jocul Asigurării; Jocul Lașului; Jocul Dictatorului; Jocul Miriapodului; Jocul Punctului Focal; Jocul Telefonului; Jocul Ultimatumului; Jocul Încrederii; Paradoxul Lanțului de Magazine; standul vânzătorului de înghețată
Jocurile Olimpice de Iarnă (1994), 151
Johnson, Samuel, 202, 208
judecată, 233-234, 239, 422
junta chiliană (1973-1980), 374
jurați, 71, 375, 383-384, 394, 401
justiție de tranziție, 17, 397-398, 417

Kahneman, Daniel, 81, 143, 217, 222, 245, 420
Kant, Immanuel
maxima de a nu fuma, 90
kantianism de zi cu zi, 108-109
Keillor, Garrison, 362
Keynes, John Maynard, 430
Kissinger, Henry, 311

La Bruyère, Jean de, 13, 54, 95, 98, 327
la Fayette, Doamna de, 232
La Rochefoucauld, François de, 12-13, 55, 110-111, 150, 159, 327
Lake Wobegon Days (Keillor), 362
Lameth, Alexandre, 391
Lanjuinais, Jean-Denis, 390

BCU CENTRAL UNIVERSITY LIBRARY

Lefebvre, Georges, 112, 315-317, 361, 413
lege
a cererii, 46
a lui Engel, 46
a gravitației, 38, 45-46
slabă, 46
vezi și mecanism
Leibniz, Gottfried Wilhelm, 237, 351
Levinson, Sam, 35
Liaisons dangereuses, Les, 339
lipsă de interes, 392
Livingston, Robert, 86
loterii, preferințe pentru, 192
Lucien Leuwen (Stendhal), 232, 240

Macaulay, Thomas, 34
Madison, James, 70, 95, 169-170, 174, 377, 379, 397
Mafia, 176, 185
mândrie, 314
mânie, 335, 340, 429
Mansfield Park (Austen), 244-245
mântuire, 16, 65, 85, 92, 121, 134, 181, 237, 415
Mao Zedong, 233, 235, 408, 410
Marbury *v.* Madison (1854), 397
Marea Teamă (1789), 355-356
Marx, Karl, 70, 101, 104, 153, 164, 172, 289-290, 296, 309, 420-421
Masa Rotundă (Polonia, 1989), 392
Masacrul din Noaptea Sfântului Bartolomeu, 69
Mason, George, 85, 105, 314, 403, 405-406
Maugham, Somerset, 72
maxime, 280
vezi proverbe și maxime
măgarul lui Buridan, 218
mărturiile făcute pe patul de moarte
în Anglia secolului al XIX-lea, 69
în India secolului al XIX-lea, 72
mecanică cuantică, 38
mecanisme
atomice, 52
și moleculare, 52
vezi și adaptarea hedonică; anticonformism; autosugestie; autoînșelare; aversiune față de risc; cascadă informațională; completare a golurilor; conformism; destrămare; disonanța cognitivă; disponibilității; efectul bulgărelui de zăpadă; efectul de certitudine; efectul de compensare; efectul de contrast; efectulde disjuncție; efectul de halo; efectul de răspândire; efectul Giffen; efectul Othello; efectul Veblen; efectul împroprietăririi; emoție, timp de înjumătățire al; euristica reprezentativității; euristica fals efect al consensului; fructul oprit; gruparea alegerilor; gândirea antiiluzorie; gândire iluzorie; gândire magică; ignoranță pluralistă; mecanismul candidatului inferior; mecanismul alierii cu cel mai puternic; preferințe adaptive, contraadaptive; punctul focal; raționalizare; reactanță; reciprocitate; selecție; sindromul fratelui mai mare; sindromul fratelui mai mic; strugurii acri; transmutație; variație
mecanismul de susținere a candidatului inferior, 52
Medeea, 122, 220, 239
Meegeren, Han van, 79
memorii (Talleyrand), 309
Mendès-France, Pierre, 283
Merleau-Ponty, Maurice, 294
metoda ipotetico-deductivă, 30
Middlemarch (Eliot), 242
milă, 149, 152
militanți (activiști), 283, 369
Miller, Arthur, 31
miopia deciziei, 216, 226, 416
miopie, 82, 114, 204
vezi și deprecierea timpului
mișcări de Rezistență, 325
Mitterrand, François, 328
Mockus, Antanas, 108

model aditiv, 55, 57, 59
modelul credință-dorință al acțiunii, 75, 429
monitorizare, 399, 401
Montaigne, Michel de
despre altruism, 102
despre Aristodemos, 92
despre Diagoras, 131
despre emoții, 151, 156
despre fapte, 92, 102
despre ignoranță, 129
despre ipocrizie, 72
despre încredere, 325, 330
despre onoare, 213, 340
despre răzbunare, 340
despre virtute, 110
despre zvon, 356-357
Morgenstern, Oskar, 192, 206
Morris, guvernatorul, 105, 314
motiv, dorința de a acționa dintr-un m., 214, 218
motivații
altruiste, 100-101, 107, 109, 256
abordări ale m., 90
în acțiunea colectivă, 370
consecinționiste, 87
continuum de m., 82
ierarhia, 414
inteligibilitatea, 65
meta-m., 95
non-consecinționiste, 87, 90
puterea m., 384
care țin cont de celălalt, 107-108
viscerale, 98, 230
Mounier, Jean-Joseph, 391
mutație, 28, 41, 265
mutilarea organelor genitale, 205
muzică clasică, 167

Nagy, Imre, 309
Napoleon al III-lea, 309, 357
Napoleon, 103-104, 328, 355, 357-358, 398
natură, starea de, 340
Naturwissenschaften, 61
Navarre, Henri de, 68
vezi și Henric al IV-lea
necesitare, 36, 41, 43
negociere
depresia ca instrument pentru n., 249, 276
a salariilor, 312
negustori de diamante, 327, 331
neîncredere, 324-325, 327, 330, 343, 397, 402-403
Nemours, Duce de, în *Prințesa de Clèves*, 232
neocortex, 251
nerăbdare, 90, 153-159
Neumann, John von, 192, 206
Newton, Isaac, 38, 249, 351
Nixon, Richard, 18, 311, 418
norme
condiționate, 108
și conformism, 338
ale consumului de alcool, 343
cvasi-morale, 108, 373
de etichetă, 346-347
și externalități, 337
morale, 108, 222, 275, 335, 373
necondiționate, 333
ale onoarei, 84, 340
privitoare la bacșiș, 344, 346
care reglementează utilizarea banilor, 342
sociale, 104, 108, 111, 222, 248, 275, 335-338, 341-342, 345, 371
norocul începătorului, 261
Nu încape îndoială (Andersen), 355
nucleotide, 264-268
nucleul caudat, 255

obsesie, 429
obstrucție (politică), 407
Odette, în *A la recherche du temps perdu*, 142-143
Odiseea (Homer), 232
omucideri, 24
onoare
coduri ale o., 339-341, 347
concepții despre o., 103
operațiuni de extorcare, 261
oportunism, 68, 324-325
oportuniști, 283-284, 326

BOSTON CENTRAL UNIVERSITY LIBRARY

oportunități
explicații bazate pe o., 165
optimalitatea rezultatului, 188
organizații, 233, 285, 363, 369, 396-399, 409
Originea speciilor (Darwin), 277
Othello, 136, 156-157, 240
oxitocină, 253-254

paracetamol, 168-169
paradox
al apei reci, 212
al bonusului în acțiuni, 212
al cardului de credit, 211
al clubului de Crăciun, 211
al dentistului, 212
al lanțului de magazine, 319, 321
al tunsului gazonului, 211
parazitul, 306, 369
pariul lui Pascal, 121, 125, 134, 136
pariuri, 215, 427
Parker, Charlie, 178
partener de viață, alegerea unui, 115
Pascal, Blaise, 13
paternalism al statului, 234
pedeapsă altruistă, 271
pedeapsa cu moartea, 423
Pélissier, generalul Aimable Jean-Jacques, 309
perdantul ia tot, 95
Pericle, 377
Persuasiune (Austen), 145
pescuit în exces, 291
Piața Comună, 354
piețe financiare, 359, 361
planificare
coerentă, 119
a unui război, 200, 202
plănuire împotriva sinelui, 229
pleiotropie, 269, 277
plictis, 159, 241
pofte viscerale, 83
politica de eugenie, 278
politica scaunului liber, 376, 381
Political Tactics (Bentham), 394, 420
Polonius, 237
Postmodern Pooh (Crew), 412
postmodernism, 412
preangajament, 204, 224, 230, 232, 234, 311-312, 324, 425
predestinare, 64-65, 133
predicție, 40-41, 52, 87, 183, 275, 424, 429
preferințe contraadaptive, 49
preferințe
adaptive, 49
agregarea, 382, 384, 388
ciclice, 287, 387
complete, 192
intensitatea, 386
inversarea, 124-125, 155, 229-230
lexicografice, 190-191
reprezentare greșită, *vezi* reprezentare greșită a preferințelor
cu un singur vârf, 386
sociale, 107
tranzitive, 189, 191-192, 386
presiune, 65, 93-94, 99, 129, 153-154, 156-157, 159, 197, 221, 267, 276, 344, 348, 352, 403, 423
prezența la vot, 24-26, 219, 383
Price, Fanny, în *Mansfield Park*, 244
primate, 115, 124
Principesa de Clèves (Doamna de la Fayette), 232
principiul echilibrului instituțional, 402
principiul plăcerii, 86, 90, 114, 138, 178
principiul rațiunii insuficiente, 427
principiul realității, 86, 138
privare relativă, 66
Privește, înger, către casă (Wolfe), 239
probabilitate
legile, 261
evaluare a p., 130
problema director-agent, soluții la, 400
Procesul de la Nürnberg, 398
promisiuni, credibilitatea p., 391
proteină, 264, 266-267, 398
Proust, Marcel
despre autoînșelare, 142
despre căutarea fericirii, 92
despre gândirea magică, 133
despre gelozie, 251, 327

DACIA CENTRAL UNIVERSITY LIBRARY

despre virtute, 102
Prouvost, Jean, 92
proverbe și maxime
Absența micșorează pasiunile moderate, dar le stârnește pe cele mari, așa cum vântul stinge o lumânare, dar aţâţă un foc, 55
Aşchia nu sare departe de trunchi, 48
Cel care minte e și hoț, 175
Cine-n pripă se însoară toată viața-i e amară, 155
Cine comite o faptă rea nu poate ierta, 252
Cine e prins cu mâța-n sac o dată va inspira mereu neîncredere, 175
Cine fură azi un ou mâine va fura un bou, 175
Cine se aseamănă se adună, 48
Cine spune o minciună va spune o sută, 175
Cine știe să păstreze un secret mic știe să păstreze și unul mare, 175
Credem cu ușurință ceea ce sperăm și ceea ce ne inspiră teamă, 358
Dacă nu e stricat, nu ai ce să repari, 280
Doi ochi pentru un ochi, 416
E nevoie de o singură oaie rea pentru a distruge o turmă, 52
E scump să fii sărac, 50
Extremele se atrag, 48
Fericirea trecută sporește nefericirea prezentă, 149
Frica este adesea mai mare decât pericolul în sine, 52
Frica sporește pericolul, 52
Graba strică treaba, 48
În fiecare turmă există o oaie neagră, 52
Întârzierea în răzbunare face lovitura mai puternică, 213
M-ai păcălit o dată, să-ți fie rușine; m-ai păcălit de două ori, să-mi fie rușine, 329
Mai răruț, dar mai drăguț, 55
Natura propune; omul dispune, 277
Nevoia este mama tuturor invențiilor, 171
Nu schimba niciodată echipa învingătoare, 280
Nu uita niciodată un serviciu care ți se face, uită-l imediat pe cel făcut de tine, 370
Oamenii sunt foarte vanitoși și detestă, dintre toate lucrurile, să fie considerați astfel, 95
Ochii care nu se văd se uită, 48
O scurtă absență poate face mult bine, 55
Părinte rău, fiu risipitor, 48
Pe aceia pe care îi rănesc, pe aceia îi și urăsc, 158
Prietenia înseamnă să uiți ce dăruiești și să ții minte ce primești, 329
Răzbunarea se servește cel mai bine rece, 213
Reamintirea pericolelor trecute este plăcută, 48
Reamintirea unui necaz înseamnă a-l mai trăi o dată, 48
Si jeunesse savait, si vieillesse pouvait (Dacă tinerii ar ști, dacă bătrânii ar putea), 172
Un ochi pentru un ochi, 416
Un rău căruia nu i se răspunde cu un rău mai mare nu a fost răzbunat, 153
Vârsta a doua este cea în care ești suficient de bătrân pentru a ști că nu e bine să faci un lucru și suficient de tânăr ca totuși să îl faci, 172
Virtutea nu se cunoaște pe sine, 78
Zvonurile sunt îndeobște mincinoase, 140
pruncuciderea pentru a favoriza un anumit sex, 278
psihanaliză kleiniană, 412
psihanaliză lacaniană, 412
psihologie populară, 715
psihologie a tiraniei, 50, 88, 155, 420
psihopați, 179
punct de maximum local, 114, 238, 267-268, 279-280
puncte focale competitoare, 430
Putnam, Robert, 421

BCU CENTRAL UNIVERSITY LIBRARY

Raffles, în *Middlemarch*, 242
ranchiună, 53, 314
Randolph, Edmund, 403
raționalitate
  atracție normativă, 207
  de tipul „ca și cum", 283, 428
  a credințelor, 205
  și egoism, 189
  principii canonice ale r., 208
  procesuală, 188
raționalizare, 404
Rawls, John, 64
răpiri
  în Columbia, 69
  în Costa Rica, 71
  în Filipine, 68
răscoale țărănești, 100, 112, 172
Războiul civil din Franța (Marx), 70
Războiul din Crimeea, 309
răzbunare
  în *Hamlet*, 17, 417
reactanță, 51
Reagan, Ronald, 23, 168
Rebel fără cauză, 306
reciprocitate
  diadică, 106
  multilaterală, 106
  negativă, 106
Reculer pour mieux sauter, 115
recunoștință, 54, 148, 152-153
reducționism
  brut, 248
  prematur, 248
  speculativ, 248
reformatori, 283-284
Regele Lear, 96
regresia către medie, 132
regret, 150, 158, 232
regimul țarist (Rusia), 408
reguli private, 90, 227, 235, 324-325
Reinhardt, Django, 178, 185
Rênal, doamna de, în *Roșu și negru*, 80
Rênal, domnul de, în *Roșu și negru*, 240
replicare, eroarea de, 279
reprezentare greșită
  a motivațiilor, 70
  a preferințelor, 384
reprimarea amintirilor, teoria (Freud), 252
reputație, 28, 91, 107, 311, 315, 328, 345, 390
Retorica (Aristotel), 149
Richardson, Lewis, 425
risc
  aversiune față de r., 118, 193-194, 196, 209, 254, 391, 416-417
roata ruletei, 217
Rocard, Michel, 283
Roșu și negru (Stendhal), 80, 240
Rousseau, Jean-Jacques, 304, 375, 383
Ruanda, 398
Rubinstein, Ariel, 426
ruleta rusească, 217
Rumsfeld, Donald, 128
Rush, Benjamin, 83
rușine, 16, 83, 93-94, 102-104, 107, 145, 147-148, 152-153, 155, 157-158, 175, 180, 329, 331, 334-335, 339, 345, 347, 353, 371, 415-416

sadism, 90, 177
Saint-Loup, Robert, în *A la recherche du temps perdu*, 133
sancțiuni, 111, 305, 333-337, 344, 363, 377, 394
Sassoulitch, Vera, scrisoarea lui Marx către, 70
satisfacere (*satisficing*), 284, 430
satisfacere prematură, 230
Schadenfreude, 256, 259
Schelling, Thomas, 235, 312, 366, 395, 420
Schumpeter, Joseph, 413, 420
Schwärmerei, 72
sclavie, 170, 172
scrofuloză, 77, 354
scuipatul în locuri publice, 337
scutire de recrutare, 295
Segrè, Emilio, 33
Ségur, Joseph-Alexandre Pierre, 339
selecție
  efectul de s., 35

BUCHAREST CENTRAL UNIVERSITY LIBRARY

eroarea de s., 131
de grup, 271-274, 276, 417
intenționată, 274-275, 277-278
prin înrudire, 270-271
naturală, 275
Selten, Reinhard, 312, 426
semnale, 252, 272, 327-328, 331, 338, 378
vezi și semnale de alarmă
semnale de alarmă, 270-271
semne, 91, 129-130, 147, 170, 237, 240, 258, 327-328, 331, 337
Seneca, 12, 95, 153, 155, 158, 221, 232
sens, căutarea, 276
sentimentalitate, 78, 81
Sentențele (Publilius Syrus), 149
Sevastopol, 309
Sfântul Pavel, 98
Shuttleworth, James Kay, 399
Sierra Leone, 398
simonie, 92
simpatie, 149
sindicate, 168, 295, 312, 369
sindromul fratelui mai mare, 419
vezi și sindromul fratelui mai mic
sindromul fratelui mai mic, 218, 293, 295, 322, 418
vezi și sindromul fratelui mai mare
sine, 86
sinucigași, teroriști s., 67, 73, 78
sistem de juranți, 401
sistemul de patente, 397
slăbiciunea voinței, 98, 122, 124-125, 204, 236-237
Smith, Adam, 223, 289-290, 384-385, 397
societatea greacă, cultură a rușinii, 159
Socrate, 123
Sokal, Alan, 412
sondaje de opinie, 424
Sorel, Julien, în *Roșu și negru*, 80, 240
Spania lui Franco, 408
spărgători de grevă, 338
spartani, 92-93
speranță, 149-150, 182, 355, 358-359
St. Kilda (Hebride), 34
Stalin, 374
„standard clar", 308
vezi și puncte focale
standul vânzătorului de înghețată, 301
Ste Croix, G.E.M. de, 413
Steagul Viu, 362-363
Stendhal (Marie-Henri Beyle), 12-13, 145, 150, 156, 232, 240-243
Stephen, James Fitzjames, 36, 73, 223, 408, 413
Stockton, John Potter, 234, 406
stoicism, 151
strategia declanșării răspunsului cumplit, 369
structuralism, 164
struguri acri, 173
Sundt, Eilert, 278-279
supradeterminare cauzală, 36
supraeu, 86, 90, 227
supraviețuire diferențiată, 263
surpriză, 66, 143, 151, 238, 251, 253
Swann, în *A la recherche du temps perdu*, 142-143, 178
Syrus, Publilius, 54, 149

șef de echipă, 398
șofer de taxi, 328, 345
Șoimul maltez, 328
șomaj
și războaiele de agresiune, 44-45
știință
cantitativă, 412
criterii pentru ș., 411
socială calitativă, 412, 414
socială slabă, 412
științele naturii, 417, 431

talamus, 251
Talleyrand, Charles Maurice de, 309, 328
Tartuffe, 72
Tarzan din neamul maimuțelor (Burroughs), 313
Tata face întotdeauna bine ceea ce face (Andersen), 287
Teagarden, Jack, 79
tendințe de răspuns contingente, 184

tensiune, rezolvare a, 417
tentație
  preangajament împotriva t., 230
teorema juraților, 394
teorema lui Bayes, 198
teoria alegerii raționale
  alternative la t.a.r., 214
  limitări ale t.a.r., 200
  natura subiectivă a t.a.r., 187
  vezi și acțiune
teoria jocurilor, 19, 182, 302, 310, 312-313
teoria perspectivei, 153, 216-217, 427
teoria postcolonială, 412
teorii care funcționează de sus în jos (în economie), 105
terorism, 66
Thatcher, Margaret, 168
Thompson, d'Arcy Wentworth, 411
Tocqueville, Alexis de
  despre *Ancien Régime*, 407
  despre căsătorie, 35
  despre conformism, 348
  despre *corvée*, 408
  despre democrație în America, 13, 35, 170
  despre egoism, 324
  despre etalarea averii, 343
  despre Napoleon al III-lea, 309
  și religie, 172
  despre Revoluția Franceză, 36
a trage concluzii pripite, 258
tragedia bunurilor publice, 291, 337
tragedie, 242
transmutație, 105, 112, 157
Tratatul de la Versailles, 23
trăsături de caracter, 175, 177-179
trăsături de personalitate, 179
tristețe, 158
Tucidide, 310, 384
Tversky, Amos, 81, 143, 217, 245, 331, 420
Twain, Mark, 232

țintirea unui obiect în mișcare, 199

Ulise, 232
Uniunea Europeană, 313, 424
ură, 50, 352
urcarea pantei, 114, 269
utilitarism clasic, 386
utilitariști
  desăvârșiți, 370
  dezinteresați, 370
utilitate
  așteptată, 75
  cardinală, 191-194, 196
  diminuarea utilității, 194
  intrinsecă, 195-196
  liniară în probabilități, 217
  marginală, 105, 195, 426
  ordinală, 189

valori
  compensații între valori, 77
Vânătoarea de Cerbi, 302-303, 305
variantă, 24-25, 165
variație
  intenționată, 277, 280
  neintenționată, 278
Vechiul regim (Tocqueville), 351
Vermeer, Johannes, 79
Verne, Jules, 243
Verstehen, 61, 73
veto, 389, 390-391
Veyne, Paul, 59, 99, 413, 420
viața de apoi, 16, 78, 92, 121, 414
Viața lui Johnson (Boswell), 34
vicii, 175, 295
Vietnam, 176, 185, 212, 390, 392
vină, 16, 34, 54, 75, 84, 97, 147-148, 152-153, 158-159, 175, 221, 258, 331, 415
viol, 16, 221, 252, 415
violența la televizor, 48, 52
virtute, 78, 101, 102-103, 110, 121, 309
Voltaire, François-Marie Arouet de, 145, 295
vot
  adunare, 380
  prin apel nominal, 384, 419
  cu cvorum, 380
  drept electoral, 380
  majoritate, 380
  paradox al v., 210

BCU IASI CENTRAL UNIVERSITY LIBRARY

popular, 380
prin ridicarea mâinii, 380
prin ridicatul în picioare, 380
secret, în adunare, 380
cu strigare, 380
vrăjitoare, 75, 339
vrăjitorie, 196, 339, 347

Washington, George, 103, 111
Weber, Max, 61, 65, 73, 134, 188, 205-206, 413, 420
Whitehead, Alfred, 411
Wilde, Oscar, 78
World Trade Center, atacul asupra (2001), 24
vezi și 11 septembrie 2001

Yawning Heights, The (Zinoviev), 413

Zeckhauser, Richard, 217, 259
Zidane, Zinedine, 83
Zinoviev, Alexandr, 413
Zola, Emile, 178, 240
zvonuri, 69, 71, 140, 332, 355-359, 361

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

B.C.U. „M. EMINESCU" IAȘI

COMPORTAMENTUL SOCIAL

Fundamentele explicației în științele sociale

Trăim cu impresia că luăm decizii conștient și rațional, pe baza unor deliberări; în realitate, pentru noi decid emoțiile, normele, prejudecățile, voința, temerile sau frustrările. *Comportamentul social* cercetează mecanismele profunde ale acțiunii oamenilor, mobilizând întregul inventar al științelor sociale. Jon Elster creează o structură de analiză și interpretare a unor probleme aparent paradoxale, învățându-l pe cititor cum să găsească răspunsul la întrebări de genul:

De ce, uneori, ni se schimbă preferințele prin simpla trecere a timpului?

De ce pedepsele, în loc să reducă rata de incidență a comportamentului pe care îl vizează, mai degrabă o sporesc?

De ce rata sinuciderilor scade atunci când medicamentele periculoase se vând în folii, față de vânzarea lor în flacoane?

De ce este mai puțin probabil ca oamenii săraci să emigreze?

De ce sunt oamenii dispuși să investească în proiectele altora, în condițiile în care cei din urmă păstrează întregul profit?

De ce trag oamenii concluzii care nu sunt susținute de dovezi?

De ce își exprimă unii oameni opțiunea de a vota chiar și atunci când aceasta nu poate produce niciun efect în privința rezultatului de la urne?

De ce firmele investesc în stocuri masive chiar și atunci când nu anticipează o întrerupere a producției?

Jon Elster (născut la Oslo în 1940) este Profesor la Collège de France, Catedra de Raționalitate și Științe Sociale. A predat la numeroase universități de prestigiu din Statele Unite și Europa. Printre cele mai recente cărți ale sale se numără *Closing the Books: Transitional Justice in Historical Perspective* (2004) și *Reason and Rationality* (2009).

www.all.ro
ISBN 978-606-587-068-0
9 786065 870680